J. Steroid Biochem. Molec. Biol. Vol. 43, No. 8, pp. 961–971, 1992
Printed in Great Britain.

0960-0760/92 $5.00 + 0.00

ANALYSIS OF THE DUPLICATED HUMAN C4/P450c21/X GENE CLUSTER

WALTER L. MILLER,[1,2]* STEPHEN E. GITELMAN,[1] JAMES BRISTOW[1] and YVES MOREL[3]

[1]Department of Pediatrics and [2]Metabolic Research Unit, University of California San Francisco, CA 94143-0978, U.S.A. and [3]INSERM U 329 and Department of Pediatrics, Université de Lyon and Hopital Debrousse, Lyon Cedex 05, France

Summary—Gene duplications, deletions and rearrangements occur with an unusually high frequency in the region of the P450c21 genes encoding 21-hydroxylase. In the human genome, the locus contains at least 6 genes, oriented 5′ C4A, P450c21A, XA, C4B, P450c21B, XB 3′. Sequence analysis of the XA gene, of the 5′ flanking DNA of the C4A gene, and of part of the XB gene revealed that this gene cluster was duplicated by nonhomologous recombination at a CAAG tetranucleotide. The location of this duplication suggests that it may have occurred after mammalian speciation. The XA gene is abundantly expressed in the human adrenal as a stable 2.6 kb RNA, but it is not known if that RNA serves a biological function. Knowledge of the anatomy of the XA gene facilitates genetic analysis of disease-causing lesions in the P450c21B gene. Southern blotting data show that about 76% of disordered P450c21B alleles bear gene microconversions that resemble point mutations; the remaining alleles are equally distributed between gene deletions and large gene conversions.

INTRODUCTION

Gene duplications play a major role in vertebrate evolution. While one gene copy performs the original function, the redundant gene copy is released from selective pressure, permitting it to accumulate mutations that may lead to new gene functions (for reviews, see [1, 2]). The duplicated gene may also become surrounded by different flanking DNA, possibly altering its transcriptional regulation (for examples see [3, 4]).

Gene duplications are especially common in the human leukocyte antigen (HLA) compatibility locus on chromsome 6p21. The Class I and II regions of the HLA locus contain multigene families encoding cell surace recognition molecules. The Class III region of about 1000 kb lies between these two regions [5, 6] and contains many duplicated genes with diverse functions, including tumor necrosis factors α and β [5–7]; heat shock protein HSP70 [8]; complement cascade factors B and C2, C4A and C4B [9]; and the adrenal steroid 21-hydroxylase genes P450c21A and B [10, 11]. Many other genes of unknown function have also been identified in this region [12–15].

The Class III region has been studied intensively because of its relationship to human disease. The best characterized portion involves the tandemly duplicated C4/P450c21 gene cluster. 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 [16, 17]. The P450c21 genes (formally termed CYP21 [18]) 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, indicating frequent genetic recombination in this locus (for review and references, see [19, 20]).

The duplication of the C4/P450c21 gene cluster has been thought to predate mammalian speciation, as mice [21] and cattle [22, 23] 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 [24–26]. In mice, the P450c21A gene is functional while P450c21B is a pseudogene bearing a single large deletion [27, 28]. In cattle, both P450c21 genes are

Proceedings of the First International Symposium on A Molecular View of Steroid Biosynthesis and Metabolism, Jerusalem, Israel, 14–17 October 1991.
*To whom correspondence should be addressed.

functional [22, 29, 30]. 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 [31], and it has a 5′ flanking DNA insertion that confers androgen responsiveness [32]. Also, 40 kb of DNA separates the 3′ end of the mouse P450c21A gene from the 5′ end of the mouse C4B gene [33], while only 6.5 kb separates the corresponding human genes. Finally, pigs [34], whales [35], Syrian hamsters and possibly guinea pigs [36] appear to have only one P450c21 gene, suggesting this locus may have duplicated independently in cattle, mice, and people after mammalian speciation (for review see [37]).

We recently discovered an additional tandemly duplicated gene pair in the human C4/P450c21 locus [38]. These genes, operationally termed XA and XB, overlap the last exon of the P450c21 A and B genes, respectively, and are encoded by the opposite strand of DNA. Cloning and sequencing of a 2.7 kb X cDNA fragment [38] showed that the encoded protein has juxtaposed fibrinogen and fibronectin type III-like domains [39, 40], thus resembling the extracellular matrix protein tenascin (reviewed in [41]). P450c21 gene deletions causing 21-hydroxylase deficiency never extend into the XB gene, but deletions of XA are found in 14% of human chromosomes [19, 20, 42]. This 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 [38].

DUPLICATION OF THE HUMAN C4/P450c21/GENE X LOCUS

We recently determined that the human C4/P450c21/X locus was duplicated by a non-homologous recombination that probably occurred after mammalian speciation. The mechanism and consequences of this duplication are summarized in Fig. 1, and have been described in detail recently [43]. The arrangement of the duplicated C4/P450c21/X gene cluster indicated that an unequal crossover event occurred between ancestral chromosomes. The approximate limits of the duplication were known from previous mapping of this region [44, 45] suggesting a duplication of about 30 kb.

Various phage clones encompassing the predicted gene duplication boundaries were isolated (Fig. 1) and several thousand bases of each of the three duplication boundaries were sequenced. The upstream duplication boundary (I), is where the sequences of the 5′ flanking DNA of the C4A and C4B genes become dissimilar. The junctional boundary of the duplication (II), is where XA sequences are joined to C4B sequences. The downstream duplication boundary (III), is where the XB gene sequence diverges from that of the XA gene. Homology matrix analyses of DNA from regions I to III showed no regions of substantial similarity, suggesting that the duplication did not arise by homologous recombination at a repetitive element. However, such matrix analysis of regions I and II and of II and III (Fig. 2) showed substantial similarity in the regions predicted by the model in Fig. 1. The 5′ flanking DNAs of C4A and C4B (3′ halves of I and II) align exactly, and the corresponding portions of XA and XB (5′ halves of II and III) align exactly. Base-by-base alignment of these regions shows >90% sequence identity between regions I and II and II and III on each side of the duplication boundary, which is identified as a

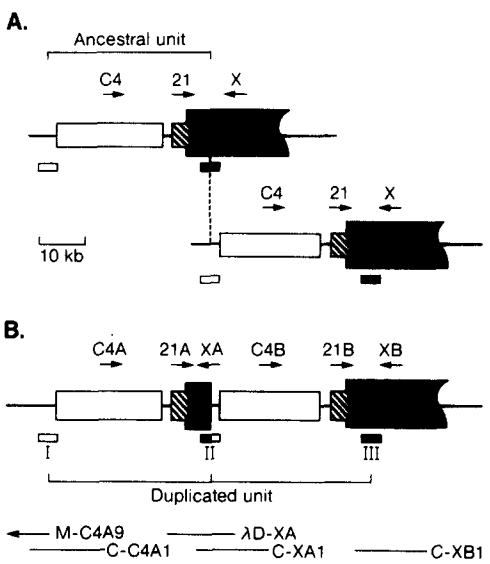

Fig. 1. (A) A model for duplication of the C4/P450c21/X locus, with unequal crossover between ancestral chromosomes. The C4 gene is represented by the open boxes, P450c21 by the hatched boxes, and X by the closed boxes; the arrows denote the transcriptional orientation. The smaller open and closed bars 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 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.

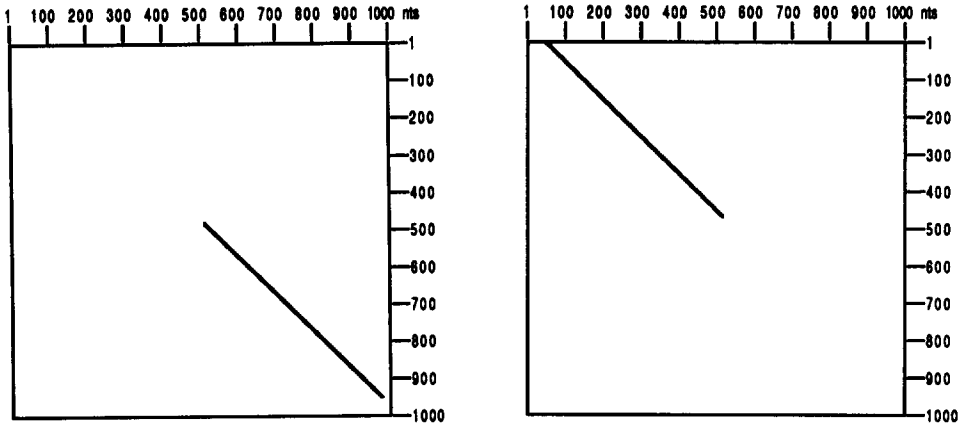

Fig. 2. Dot matrix sequence comparisons of the 1 kb regions denoted I, II, and III in Fig. 1 was performed using the DNA Inspector (Textco, W. Lebanon, N.H.). The *left* panel compares regions I (x-axis) with II (y-axis) and the *right* compares region III (x-axis) with II (y-axis). Each dot represents a region in which there is identity at 18 of 20 base pairs. Note that the break point within region II in the two comparisons occurs at precisely the same point. Comparison of region I with III showed no significant homology (not shown).

block of 4 bases (CAAG) found in all three sequences [43]. Such regions of "patchy homology" are found in over half of nonhomologous recombinations [46–49], such as those in the β-globin locus [50] and the *aprt* locus in Chinese hamster ovary (CHO) cells [51].

The effects of this gene duplication on the expression of nearby genes is under investigation. For example, Sargent *et al.* [13] 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 (see Fig. 2 in Ref. [43]). The one at bases −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 if either of these sites corresponded to the 3′ end of the G11 gene, and to determine if a portion of the G11 gene was duplicated with the C4/P450c21/X locus, we probed Northern blots of U937 RNA with two genomic fragments, one lying upstream and one downstream from the duplication boundary. Neither of these probes hybridized to a 1.4 kb mRNA as reported for G11 [13]. The upstream probe hybridized at approx. 10 kb and the downstream probe hybridized at 4.8 kb (Fig. 3). 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. However, the greatest impact of this gene duplication was clearly on the XA gene.

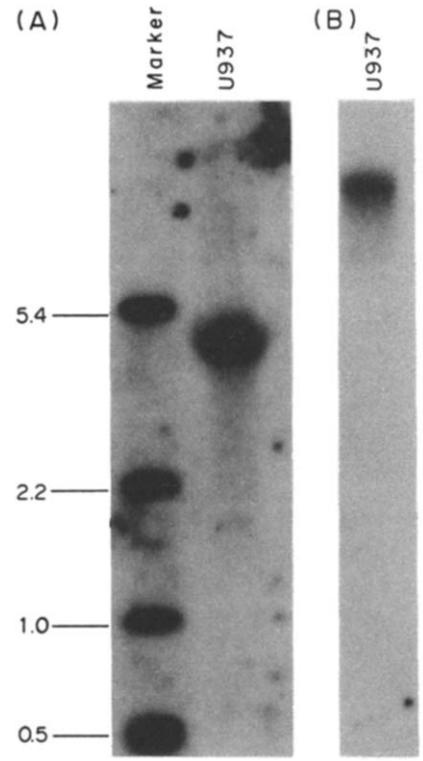

Fig. 3. Northern blot analysis of total RNA from U937 cells. (A) The blot was probed with a genomic fragment unique to the 5′ flanking DNA of the C4A gene (extending from nucleotide −2340 to the *Pst* I site at −1547; see Fig. 2 in Ref. [43]. (B) The same blot was stripped and reprobed with a 1 kb *Apa*I genomic DNA fragment common to the 5′ flanking DNA of both the C4A and C4B genes (nucleotides −1336 to −347 in Fig. 2 in Ref. [43]). The markers are bacteriophage PM2 digested with *Hin*d III.

STRUCTURE AND EXPRESSION OF GENE XA

We previously identified an expressed gene, operationally termed gene X, that overlaps the P450c21 gene. P450c21 and X lie on opposite strands of DNA and have the opposite transcriptional orientation, and overlap only at their 3′ ends (Fig. 4). To determine if only one or both of the X genes was functional we cloned and sequenced the entire XA genomic region lying between C4A and P450c21B (see Fig. 3 in Ref. [43]). Comparison to the original cDNA sequence [38] showed two differences. First, the 5′ end of the original cDNA was not found in the XA gene, and second, the XA gene had a 121 bp internal deletion (including 91 bp of putative exon) not found in the original cDNA. Thus the cDNA we reported in 1989 [38] was encoded by XB.

We identified numerous X cDNAs by screening for the presence or absence of the internal deletion [43] and identified several XA cDNAs. One of these containing 2450 bp plus a poly(A) tail was nearly full-length (Fig. 5), as it corresponded to the 2.6 kb band seen in Northern blots [43]. The XA cDNA contained the 91 bp internal deletion predicted by the gene sequence; as a result, the long open reading frame found in the XB cDNA was closed so that XA has no long open reading frame initiated by a typical ATG sequence. However, it does have a 933 base open reading frame initiated by a CTG sequence in an appropriate context of surrounding bases that might permit this CTG to be used to initiate translation [52]. Such unusual CTG translation initiators are well-described in several other systems [53–56], but we do not know if such a CTG codon is used to initiate translation of XA, or if XA is translated at all. Nevertheless, XA is expressed only in the adrenal, where its RNA is stable and about as abundant as that for XB or P450c21B [43].

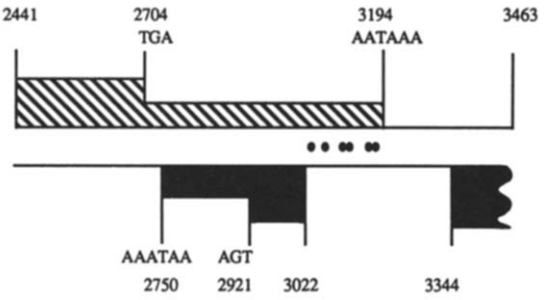

Fig. 4. Diagram of the overlapping portions of the P450c21 and X genes. The upper line shows the 3′ end of the P450c21B gene; the base numbers are from Higashi *et al.* [24]. The broader part of the hatched bar indicates the portion of exon 10 encoding P450c21 protein; the narrower portion of the hatched bar between the TGA translational stop codon and the AATAAA polyadenylation site designates the portion of exon 10 encoding the 3′ untranslated portion of P450c21 mRNA. The line extends to base 3463, the 3′ limit of the 3′ flanking DNA of the P450c21B gene sequenced by Higashi *et al.* [24]. The solid bars indicate the last two exons of the XB gene; note that all of the last exon of XB is encoded on the opposite strand of DNA from that encoding the 3′ untranslated region of P450c21B. The TGA stop codon and AATAAA polyadenylation sites for XB are shown in reverse orientation. The six dots between the two diagrams show the locations of the only base differences between the P450c21A and P450c21B genes in exon 10.

```
   1: CTCGGCAGGACACCGGGGTGTGGAAGGGGGAAGCGAGCACCTGACTCAGA
  51: CAGCGCGGGAGCTCGCAGGAGTCACGAGGCCACAGCGACTTCATTGTCTG
 101: ACTGGGCCTGGACCTATAAACTTCCCACCTCAGCCTTGGGCCAAGCCTGG
   1:                                                  L
 151: AAGATAAAAATGGAGCACCCCATGGCGCCCCTCACTCAGATTCTCCCCTG
   2: E  D  K  N  G  A  P  H  G  A  P  H  S  D  S  P  L
 201: GGCTTCTCCCACGCAGCCCCAGAAGAGGACACACCAGCCCCAGAGTTAGC
  19:  G  F  S  H  A  A  P  E  E  D  T  P  A  P  E  L  A
 251: CCCAGAGGCCCCTGAGCCTCCTGAAGAGCCCCGCCTAGGAGTGCTGACCG
  36:    P  E  A  P  E  P  P  E  E  P  R  L  G  V  L  T
 301: TGACCGACACAACCCCAGACTCCATGCGCCTCTCGTGGAGCGTGGCCCAG
  52: V  T  D  T  T  P  D  S  M  R  L  S  W  S  V  A  Q
 351: GGCCCCTTTGATTCCTTCGTGGTCCAGTATGAGGACACGAACGGGCAGCC
  69:  G  P  F  D  S  F  V  V  Q  Y  E  D  T  N  G  Q  P
 401: CCAGGCCTTGCTCGTGGACGGCGACCAGAGCAAGATCCTCATCTCAGGCC
  86:    Q  A  L  L  V  D  G  D  Q  S  K  I  L  I  S  G
 451: TGGAGCCCAGCACCCCCTACAGGTTCCTCCTCTATGGCCTCCATGAAGGC
 102: L  E  P  S  T  P  Y  R  F  L  L  Y  G  L  H  E  G
 501: AAGCGCCTGGGGCCCCTCTCAGCTGAGGGCACCACAGGGCTGGCTCCTGC
 119:  K  R  L  G  P  L  S  A  E  G  T  T  G  L  A  P  A
 551: TGGTCAGACCTCAGAGGAGTCAAGGCCCCGCCTGTCCCAGCTGTCTGTGA
 136:    G  Q  T  S  E  E  S  R  P  R  L  S  Q  L  S  V
 601: CTGACGTGACCACCAGTTCACTGAGGCTCAACTGGGAGGCCCCACCGGGG
 152: T  D  V  T  T  S  S  L  R  L  N  W  E  A  P  P  G
 651: GCCTTCGACTCCTTCCTGCTCCGCTTTGGGGTTCCATCACCAAGCACTCT
 169:  A  F  D  S  F  L  L  R  F  G  V  P  S  P  S  T  L
 701: GGAGCCGCATCCGCGTCCACTGCTGCAGCGCGAGCTGATGGTGCCGGGGA
 186:    E  P  H  P  R  P  L  L  Q  R  E  L  M  V  P  G
 751: CACGGCACTCGGCCGTGCTCCGGGACCTGCGTTCCGGGACTCTGTACAGC
 202: T  R  H  S  A  V  L  R  D  L  R  S  G  T  L  Y  S
 801: CTGACACTGTATGGGCTGCGAGGACCCCACAAGGCCGACAGCATCCAGGG
 219:  L  T  L  Y  G  L  R  G  P  H  K  A  D  S  I  Q  G
 851: AACCGCCCGCACCCTCAGCCCAGTTCTGGAGAGCCCCCGTGACCTCCAAT
 236:    T  A  R  T  L  S  P  V  L  E  S  P  R  D  L  Q
 901: TCAGTGAAATCAGGGAGACCTCAGCCAAGGTCAACTGGATGCCCCCACCA
 252: F  S  E  I  R  E  T  S  A  K  V  N  W  M  P  P  P
 951: TCCCGGGCGGACAGCTTCAAAGTCTCCTACCAGCTGGCGGACGGAGGGGA
 269:  S  R  A  D  S  F  K  V  S  Y  Q  L  A  D  G  G  E
1001: GCCTCAGAGTGTGCAGGTGGACGGCCGGGCCCGGACCCAGAAACTCCAG▼
 296:    P  Q  S  V  Q  V  D  G  R  A  R  T  Q  K  L  Q
1050: TTCCTGACGGTCCCACACAGTTGCGTGCACTGAACTTGACCGAGGGATTC
 312:  F  L  T  V  P  H  S  C  V  H  *
1100: GCCGTGCTGCACTGGAAGCCCCCCCAGAATCCTGTGGACACCTATGACGT
1150: CCAGGTCACAGCCCCTGGGGCCCCGCCTCTGCAGGCGGAGACCCCAGGCA
1200: GCGCGGTGGACTACCCCCTGCATGACCTTGTCCTCCACACCAACTACACC
1250: GCCACAGTGCGTGGCCTGCGGGGCCCCAACCTCACTTCCCCAGCCAGCAT
1300: CACCTTCACCACAGGGCTAGAGGCCCCTCGGGACTTGGAGGCCAAGGAAG
1350: TGACCCCCCGCACCGCCCTGCTCACTTGGACTGAGCCCCCAGTCCGGCCC
1400: GCAGGCTACCTGCTCAGCTTCCACACCCCTGGTGGACAGAACCAGGAGAT
1450: CCTGCTCCCAGGAGGGATCACATCTCACCAGCTCCTTGGCCTCTTTCCCT
1500: CCACCTCCTACAATGCACGGCTCCAGGCCATGTGGGGCCAGAGCCTCCTG
1550: CCGCCCGTGTCCACCTCTTTCACCACGGGTGGGCTGCGGATCCCCTTCCC
1600: CAGGGACTGCGGGGAGGAGATGCAGAACGGAGCCGGTGCCTCCAGGACCA
1650: GCACCATCTTCCTCAACGGCAACCGGGAGCGGCCCCTGAACGTGTTTTGG
1700: GACATGGAGACTGATGGGGGCGGCTGGCTGGTGTTCCAGCGCCACATGGA
1750: TGGACAGACAGACTTCTGGAGGGACTGGGAGGACTATGCCCATGGTTTTG
1800: GGAACATCTCTGGAGAGTTCTGGCTGGGCAATGAGGCCCTGCACAGCCTG
1850: ACACAGGCAGGTGACTACTCCATGCGCGTGGACCTGCGGGCTGGGGACAA
1900: GGCTGTGTTCGCCCAGTACGACTCCTTCCACGTAGACTCGGCTGCGGAGT
1950: ACTACCGCCTCCACTTGGAGGGCTACCACGGCACCGCAGGGGACTCCATG
2000: AGCTACCACAGCGGCAGTGTCTTCTCTGCCCGTGATCGGGACCCCAACAG
2050: CTTGCTCATCTCCTGCGCTGTCTCCTACCGAGGGGCCTGGTGGTACAGGA
2100: ACTGCCACTACGCCAACCTCAACGGGCTCTACGGGAGCACAGTGGACCAT
2150: CAGGGAGTGAGCTGGTACCACTGGAAGGGCTTCGAGTTCTCGGTGCCCTT
2200: CACGGAAATGAAGCTGAGACCAAGAAACTTTCGCTCCCCAGCGGGGGAG
2250: GCTGAGCTGCTGCCCACCTCTCTCGCACCCCAGTATGACTGCGAGCACT
2300: GAGGGGTCGCCCCGAGAGAAGAGCCAGGGTCCTTCACCACCCAGCCGCTG
2350: GAGGAAGCCTTCTCTGCCAGCGATCTCGCAGCACTGTGTTTACAGGGGGG
2400: AGGGGAGGGGTTCGTACGGGAGCAATAAAGGAGAAACTGAGGTACCCGGC
```

Fig. 5. Sequence of the XA cDNA. The single letter amino acid code is aligned below the middle nucleotide of each codon. The site of the 91 bp deletion is marked with an arrowhead, and the polyadenylation signal is underlined. A poly A tail follows the last nucleotide shown.

IS XA A PSEUDOGENE?

Like many newly coined words, the term "pseudogene" is used differently by different investigators. Some have used the term solely in the context of intron-less pseudogenes inserted into the genome by "retroposon" events. Others might say that a pseudogene is any gene that has lost the ability to be transcribed. Still others might say that transcription is irrelevant; that if the gene has no "function" it is a pseudogene. We shall avoid this semantic quagmire and consider what we know about XA.

First, it appears that X genes are a general mammalian feature: the available sequences of the 3′ ends of the bovine [29] and murine [28] P450c21 genes show the same array of open reading frames, stop signals and intron/exon junction sequences seen in Fig. 4. Second, the intragenic distance between C4A and P450c21B, i.e. the space into which an XA gene must fit, varies substantially among the murine [33], human [10] and bovine [29] genomes, so that these three species might have substantially different sized XA genes. Of these three, the human XA region is the smallest. Human XA could encode a protein initiated from the CTG codon discussed above; such a protein might function as a small extracellular matrix protein, as it would consist solely of $2\frac{1}{2}$ repeats similar to those of type III fibronectin [43]. Alternatively, the XA RNA may have a nontranslated function. While most functional nontranslated RNAs such as rRNA, tRNA, snRNA etc. are transcribed by RNA polymerase I or III, at least one polyadenylated, nontranslated biologically functional polII transcript has been described [57, 58]. Finally XA could have no function; it might be transcribed gratuitously and simply be a vestigial gene fragment relevant only for its structural genetics and not for its cell biology. Presently available data cannot distinguish among these three possibilities.

EVOLUTION OF THE C4/P450c21/X LOCUS

To date nothing is known about the C4/P450c21/X locus in species other than mammals. Initial studies in three mammalian orders—primates (human beings) [9–11], ungulates (cattle) [22, 23, 29, 30] and rodents, (mice) [21, 59], showed duplicated C4/P450c21 loci, leading to the conclusion that the duplication of the C4/P450c21/X locus predated mammalian speciation about 85 million years ago. However this widely held conclusion is challenged by two lines of evidence. First, the duplication boundaries in the human [43] and murine [59] loci differ by about 400 bp and occur in regions lacking sequence similarity. This suggests the duplication occurred at different places in these species, suggesting it occurred after the ancestors of each species separated. Second, Southern blotting studies suggest that several mammalian species have single C4/P450c21 clusters [34–36]; however in the absence of cloning and sequencing data, these conclusions must be interpreted with caution, as such blotting studies may not distinguish duplicated genes with identical restriction maps as in some patients with adrenal hyperplasia (see Figs 6–9, below). If indeed these various mammals do not have duplicated C4/P450c21/Gene X loci, the data still could not distinguish post-speciation duplication in some selected orders of mammals, from a pre-speciation duplication with subsequent loss of one of the duplicated pairs in other selected orders. Many studies of the human C4/P450c21/X locus clearly show that both gene duplication and gene deletion continue to be common events in this locus. If this continuing plasticity of the human locus is representative of other species, then most studies indicating single C4/P450c21 loci in some species must be viewed with even greater caution, as such plasticity would require the examination of genomic DNA from multiple individuals in each species.

EFFECTS OF THE DUPLICATION

Duplication of the human C4/P450c21/Gene X gene cluster has important genetic and clinical consequences. First, as discussed above, the duplication truncated the 5′ portion of the X gene to yield a shortened XA gene that is only 6 kb long and encodes a 2.6 kb RNA. By contrast, the XB gene is quite large and appears to encode two mRNAs of 8 and 12 kb [40, 43]. Other effects on XA are discussed above. Second, the close proximity of the C4A and C4B genes and of the P450c21A and P450c21B genes facilitates frequent genetic recombinations leading to frequent gene conversions. Such conversion between C4A and C4B [60–63] and between P450c21A and P450c21B [42, 64–70] have been well documented. The high incidence of gene conversions between the P450c21A and P450c21B loci is the most important factor in the unusual genetics of this locus (for reviews

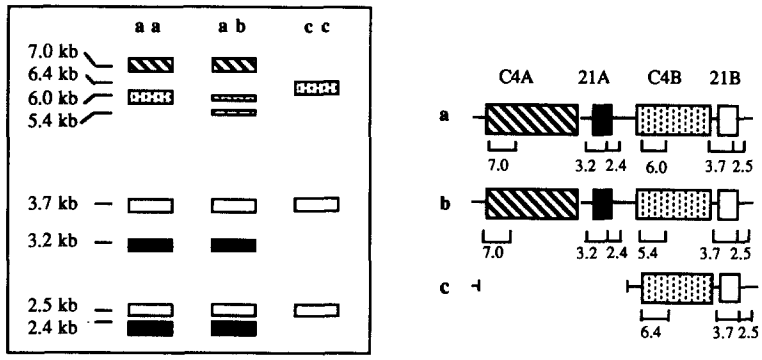

Fig. 6. Diagram of Southern blotting patterns of genomic DNA digested with *Taq* I and probed with a mixture of a 5′ C4 cDNA probe such as the 476 bp *Bam* HI/*Kpn* I fragment of clone pAT-A [73] and a genomic 3.1 kb *Bam* HI/*Eco* RI probe extending from the *Eco* RI site in the P450c21B gene to the *Bam* HI site in its 3′ flanking DNA. In the left diagram, the bands are filled-in or cross-hatched in a pattern corresponding to the gene pattern (right). The C4A gene (broad diagonal lines) is represented by a 7.0 kb band; the long C4B gene (dotted) by a 6.0 kb band; the short C4B gene (dotted) by a 5.4 kb band; and the C4B gene (dotted) by a 6.4 kb band when it is associated with a deletion of the C4A and P450c21A genes. The P450c21A pseudogene is represented by a 3.2 kb band and the 3′ end of the XA gene is represented by a 2.4 kb band (solid). The P450c21B gene is represented by a 3.7 band and the 3′ end of the XB gene is represented by a 2.5 kb band (open). Three different *Taq* I blotting patterns are diagrammed on the left corresponding to the chromosomal patterns on the right. The **a a** pattern and **a b** pattern are normal corresponding to the normal chromosomes **a** and **b**, which differ in the presence of a long (a) or short (b) C4B gene. Chromosome **c** shows deletion of C4A and the P450c21A pseudogene, characterized by the presence of a 6.4 kb band at the 5′ end of the C4B gene and absence of the 3.2 and 2.4 kb fragments. All three of these chromosomes yield a normal phenotype unassociated with disease.

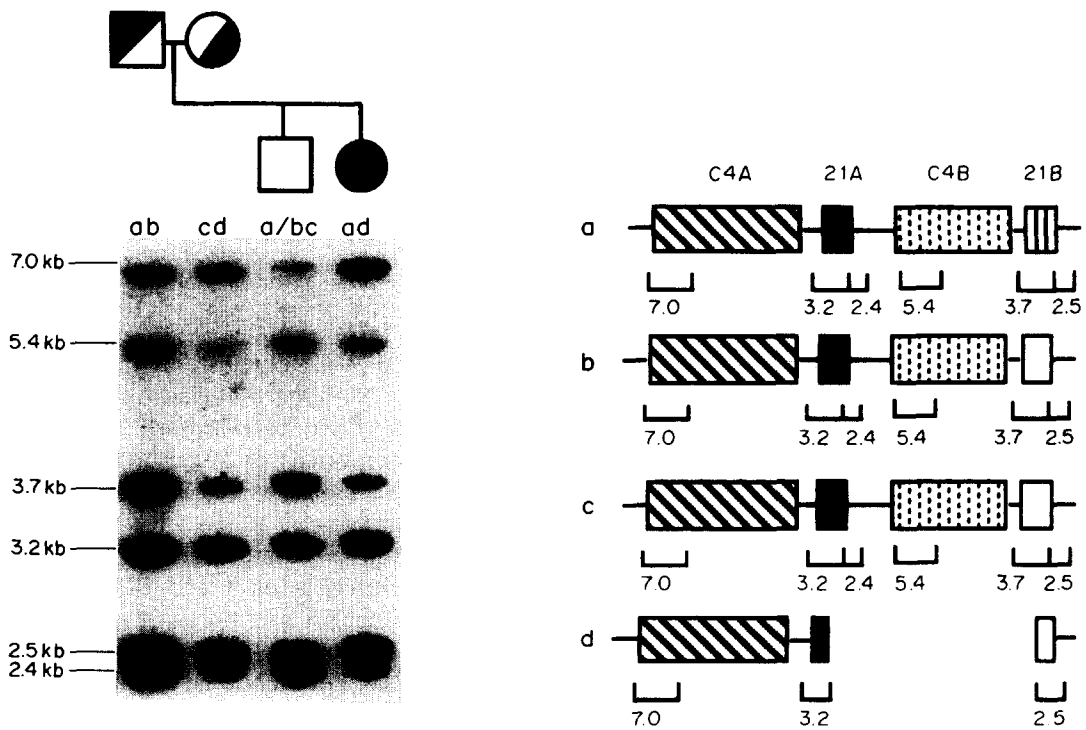

Fig. 7. Typical family study of P450c21 genes in CAH. The father carries mutant chromosome **a**, bearing a microconversion in the P450c21B gene (vertical lines); the mother carries mutant chromosome **d** bearing a 30 kb deletion extending from the middle of the P450c21A pseudogene to the middle of the P450c21B gene. The father's Southern blotting pattern is normal. The mother's pattern shows decreased intensities of the 5.4 kb band (absent C4B on **d**), of the 3.7 kb band (absent 5′ end of the P450c21B gene on **d**) and of the 2.4 kb band (absent 3′ end of XA on **d**). The unaffected son has a normal Southern blotting pattern; this study cannot determine if he received chromosome **a** or **b** from his father. The affected daughter has the same Southern blotting pattern as the mother.

see [19, 20]). Thus even though P450c21A is a pseudogene, it has an enormous effect on the genetics and pathology of the locus. While it is not yet known if genetic exchange leading to gene conversion occurs between the XA and XB genes, the extreme similarity of their sequences suggests this may well occur. If so, XA may assume substantial biological importance, irrespective of the semantic question about whether or not it is a "pseudogene"

ALTERATIONS OF THE C4/P450c21/X LOCUS IN CONGENITAL ADRENAL HYPERPLASIA (CAH)

CAH is a group of autosomal recessive disorders resulting from mutations in the genes encoding the steroidogenic enzymes. While mutations affecting each step in steroidogenesis are known, the overwhelmingly most common form of CAH involves P450c21 [71, 72]. The genetics of the C4/P450c21 locus in CAH have been studied in great detail (for reviews see [19, 20]). We have developed a simple Southern blotting tactic that discriminates the C4, P450c21, and X genes simultaneously [20]. We probe *Taq* I-digested DNA with a mixture of the 476 bp *Bam* HI/*Kpn* I fragment from the 5′ end of the full-length human C4 cDNA [73] and a P450c21/X genomic probe [42] (Figs 6 and 7). Using this 5′ C4 cDNA fragment plus the 3.1 kb *Bam* HI/*Eco* RI fragment lying immediately 3′ of the P450c21B gene provides a simple unambiguous strategy for identifying rearrangements in C4/P450c21 gene units in families with

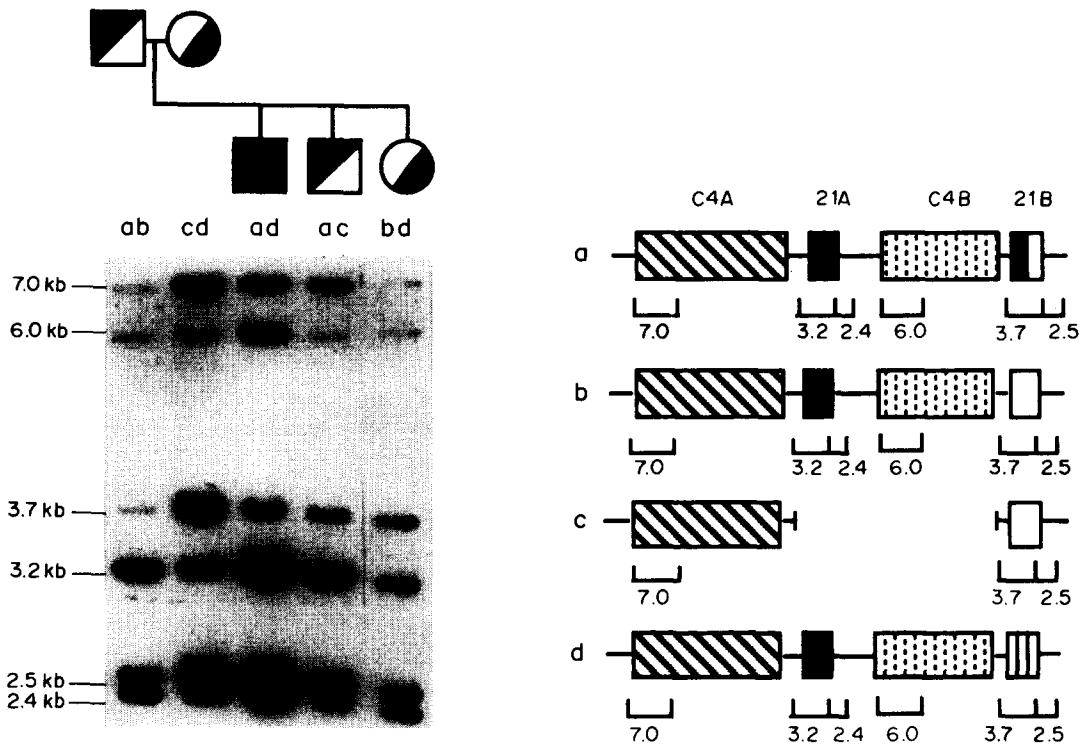

Fig. 8. Family studies are necessary in studying the genetics of CAH. The father's mutant **a** chromosome has a gene conversion in the P450c21B gene: the 7.0 and 6.0 kb C4A and C4B bands are of equal intensity; the 2.4 and 2.5 kb XA and XB bands are of equal intensity; and the 3.7 kb P450c21B band is markedly decreased (1 copy) compared to the 3.2 band (3 copies). The mother's **c** chromosome carries a normal P450c21B gene so that the mother is phenotypically normal, however, this chromosome bears deletions of the P450c21A pseudogene and the C4B gene. The mother's **d** chromosome is structurally intact, but bears a microconversion rendering the P450c21B gene nonfunctional. Thus the mother has decreased intensity of her 6.0 kb C4B band (1 copy) compared to her 7.0 kb C4A band (2 copies), and decreased intensity of her 3.2 kb P450c21A and 2.4 kb XA bands (1 copy each) compared to her 3.7 P450c21B and 2.5 kb XB bands (2 copies each). Note that the affected son carrying chromosomes **a** and **d** has at least one copy of each band and has a pattern of P450c21 bands suggesting heterozygosity for a P450c21B gene deletion or conversion (decrease of 3.7 kb compared to 3.2 kb bands); the family study disproves this by the equal intensities of the 2.4 and 2.5 kb XA and XB bands. The clinically unaffected brother carrying chromosomes **a** and **c** looks like he has a *Taq* I pattern indicating a typical heterozygote for the 30 kb deletion extending from the middle of the P450c21A to the middle of P450c21B (i.e. identical to the maternal **c d** pattern in Fig. 7). The family study shows that this pattern in the **a c** brother is due to the combination of a P450c21 gene conversion on **a** and a deletion of the P450c21A and C4B genes of **c**.

CAH. The design of this strategy was described previously [20] and is summarized in Fig. 6. The C4A gene is characterized by a 7.0 kb *Taq* I fragment, P450c21A by a 3.2 kb fragment, and XA by a 2.4 kb fragment. The C4B gene is normally polymorphic, yielding either 5.4 or 6.0 kb bands, while P450c21B is represented as a 3.7 kb band and XB as a 2.5 kb band.

Gene conversions account for the overwhelming majority of disease-causing lesions in P450c21 genes. Most P450c21 gene conversions are "micro-conversions", where only a limited region of the P450c21B gene acquires P450c21A sequences. Because these regions of microconversion are small, and because the surrounding DNA has 98% nucleotide sequence identity for the two genes, such microconversions look like point mutations, except that the mutation always changes a P450c21B sequence to a P450c21A sequence. Such microconversions account for about 75% of P450c21 mutations causing severe CAH [19, 20, 42]. Less common macroconversions change large portions of the P450c21B gene to P450c21A sequences, including the 5′ flanking DNA, so that the converted P450c21B gene is now characterized by a 3.2 kb *Taq* I band rather than a 3.7 kb band. Such macroconversions account for 10–12% of disease-causing P450c21 alleles [19, 20, 42]. The application of our Southern blotting procedure in CAH is shown in Figs 7–9. Figures 8 and 9 are especially revealing as they demonstrate the absolute requirement for family studies. This is most clearly shown in the DNA of the affected individual with the **a d** genotype in Fig. 9. This individual's DNA yields a pattern of restriction fragments that is absolutely normal, yet the family studies show that both the *a* and *d* haplotypes are abnormal. Using this tactic to study both normal chromosomes and those

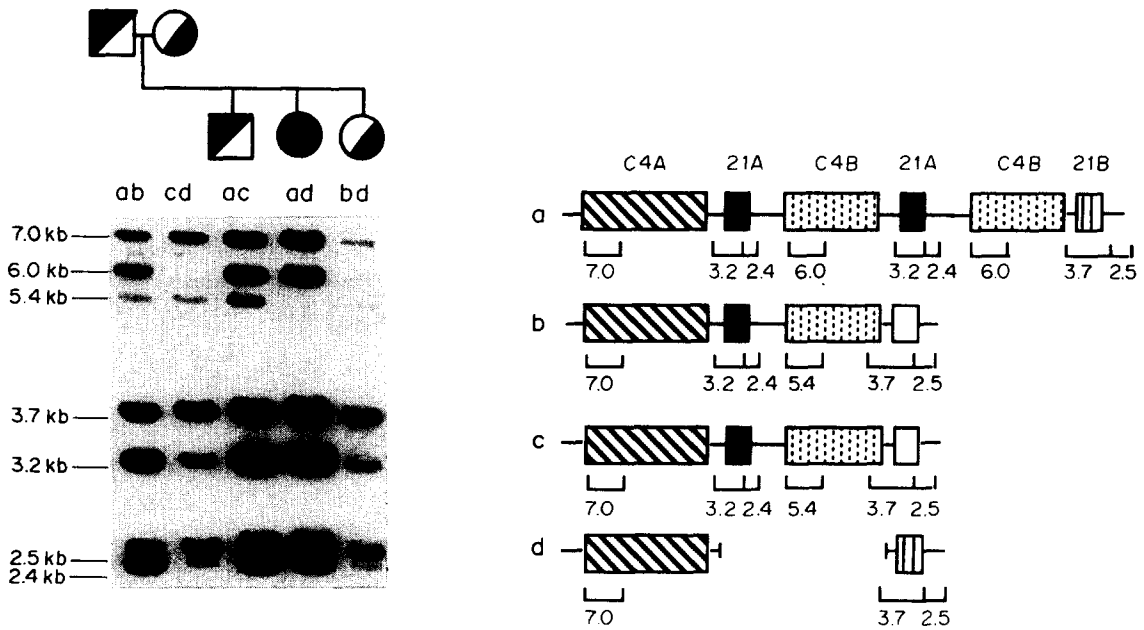

Fig. 9. Family studies show that a normal appearing pattern is not normal. Southern blots of *Taq* I-digested DNA was probed as in Figs 6–8. The affected daughter has a completely normal Southern blotting pattern suggesting a microconversion. However, the father's mutant **a** chromosome carries three C4/P450c21 gene units, with a duplication of the long C4B gene, two P450c21A pseudogenes and a P450c21B gene with a point mutation. His normal **b** chromosome is unremarkable except that it carries a short C4B gene (5.4 kb fragment). His blotting pattern shows equal intensity of the 7.0 kb C4A band (2 copies) and the 6.0 kb long C4B band (2 copies) plus a lesser intensity of the 5.4 kb short C4B band (1 copy). The presence of three P450c21A pseudogenes is indicated by the slightly greater intensities of the 3.2 kb P450c21A and 2.4 kb XA bands compared to the 3.7 kb P450c21B and 2.5 kb XB bands; however such densitometry is nonquantitative. The maternal **c** chromosome is normal, and carries a short C4B gene. Her **d** chromosome is abnormal, bearing a 30 kb deletion encompassing the entire P450c21A and C4B loci and also bearing a structurally intact P450c21B gene carrying a microconversion. The affected daughter carrying chromosomes **a** and **d** has an apparently normal pattern, with three C4/P450c21 units from paternal chromosome **a** and **d** and one from maternal chromosome **d**. Paternal chromosome **a** contributes two 3.2 kb P450c21A and 2.4 kb XA fragments; chromosomes **a** and **d** each contribute a mutated 3.7 kb P450c21B gene and its linked 2.5 kb XB fragment. The heterozygous sister having haplotype **b d** confirms the pattern, carrying the completely normal **b** chromosome and the partially deleted **d** chromosome.

Table 1. Incidence of various lesions in the C4/P450c21/X gene locus causing CAH

Mutation	Affected chromosomes: Form of CAH SW	SV	NC	Total	Incidence (%)	Normal chromosomes: Number	Incidence (%)
Classic 30 kb deletion	34	6	2	42	11.8	0	
Other deletions	5	1	1	7	2.0	0	
Macroconversions	33	2	1	36	10.1	0	
Microconversions (All)	125	76	69	274	76.1		
With a completely normal locus	102	59	27	188	53.0	254	75.4
With deleted C4A and CYP21A genes	5	0	1	6	3.4	31	9.2
With deleted C4B and CYP21A genes	2	2	2	6		20	5.9
With duplicated C4 and CYP21A genes	13	12	39	64	18.0	16	4.75
Other abnormalities	3	3	0	6	1.7	16	4.75
Total	195	85	73	355	100	337	100

The forms of CAH due to 21-hydroxylase deficiency are: SW, salt-wasting CAH; SV, simple virilizing CAH; NC, nonclassical CAH (for definitions and discussions see [19, 20].

bearing mutated alleles causing CAH we have now obtained the statistics shown in Table 1; these are substantially equivalent to those we reported earlier [19, 20, 42]. Thus knowledge of the unusual genetic anatomy of the C4/P450c21/X locus can facilitate the clinical diagnosis of genetic lesions causing CAH.

Acknowledgements—Supported by NIH Grants DK 37922 and DK 42154 to WLM, March of Dimes Grant 6-396 to WLM and NIH Fellowship DK 08325 to SEG.

REFERENCES

1. Doolittle R. F.: Similar amino acid sequences: chance or common ancestry? *Science* **214** (1981) 149–159.
2. Ohta T.: Multigene and supergene families. *Oxf. Surv. Evol. Biol.* **5** (1988) 41–65.
3. Markert C. L.: Isozymes and the regulatory structure of the genome. *Curr. Topics Biol. Med. Res.* **14** (1987) 1–17.
4. Cross M. and Renkawitz R.: Repetitive sequence involvement in the duplication and divergence of mouse lysozyme genes. *EMBO Jl* **9** (1990) 1283–1288.
5. Dunham I., Sargent C. A., Trowsdale J. and Campbell R. D.: Molecular mapping of the human major histocompatibility complex by pulsed-field gel electrophoresis. *Proc. Natn. Acad. Sci. U.S.A.* **84** (1987) 7237–7241.
6. Carroll M. C., Katzman P., Alicot E. M., Koller B. H., Geraghty D. E., Orr H. T., Strominger J. L. and Spies T.: Linkage map of the human major histocompatibility complex including the tumor necrosis factor genes. *Proc. Natn. Acad. Sci. U.S.A.* **84** (1987) 8535–8539.
7. Spies T., Morton C. C., Nedospasov S. A., Fiers W., Pious D. and Strominger J.: Genes for the tumor necrosis factors α and β are linked to the human major histocompatibility complex. *Proc. Natn. Acad. Sci. U.S.A.* **83** (1986) 8699–8702.
8. Sargent C. A., Dunham I., Trowsdale J. and Campbell R. D.: Human major histocompatibility complex contains genes for the major heat shock protein HSP 70. *Proc. Natn. Acad. Sci. U.S.A.* **86** (1989) 1968–1972.
9. Carroll M. C., Campbell R. D., Bentley D. R. and Porter R. R.: A molecular map of the human major histocompatibility complex class III region linking complement genes C4, C2, and factor B. *Nature* **307** (1984) 237–241.
10. Carroll M. C., Campbell R. D. and Porter R. R.: Mapping of steroid 21-hydroxylase genes adjacent to complement component C4 genes in HLA, the major histocompatibility complex in man. *Proc. Natn. Acad. Sci. U.S.A.* **82** (1985) 521–525.
11. White P. C., Grossberger D., Onufer B. J., Chaplin D. D., New M. I., Dupont B. and Strominger J. L.: Two genes encoding steroid 21-hydroxylase are located near the genes encoding the fourth component of complement in man. *Proc. Natn. Acad. Sci. U.S.A.* **82** (1985) 1089–1093.
12. Levi-Strauss M., Carroll M. C., Steinmetz M. and Meo T.: A previously undetected MHC gene with an unusual periodic structure. *Science* **240** (1988) 201–204.
13. Sargent C. A., Dunham I. and Campbell R. D.: Identification of multiple HTF-island associated genes in the human major histocompatibility complex class III region. *EMBO Jl* **8** (1989) 2305–2312.
14. Spies T., Blanck G., Bresnahan M., Sands J. and Strominger J. L.: A new cluster of genes within the human major histocompatibility complex. *Science* **243** (1989) 214–217.
15. Kendall E., Sargent C. A. and Campbell R. D.: Human major histocompatibility complex contains a new cluster of genes between the HLA-D and complement C4 loci. *Nucleic Acids Res.* **18** (1991) 7251–7257.
16. Porter R. R.: Complement polymorphism, the major histocompatibility complex and associated diseases: a speculation. *Molec. Biol. Med.* **1** (1983) 161–168.
17. Hauptmann G., Tappeiner G. and Schifferli J. A.: Inherited deficiency of the fourth component of human complement. *Immunodef. Rev.* **1** (1988) 3–22.
18. Nebert D., Nelson D., Coon M., Estabrook R., Feyereisenn R., Fuji-Kuriyama Y., Gonzalez F., Guengerich F., Gunsalus I., Johnson E., Loper J., Sato R., Waterman M. and Waxman D.: The P450 superfamily: update on new sequences, gene mapping, and recommended nomenclature. *DNA Cell Biol.* **10** (1991) 1–14.
19. Miller W. L. and Morel Y.: The molecular genetics of 21-hydroxylase deficiency. *A. Rev. Genet.* **23** (1989) 371–393.
20. Morel Y. and Miller W. L.: Clinical and molecular genetics of congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Adv. Hum. Genet.* **20** (1991) 1–68.
21. Amor M., Tosi M., Duponchel C., Steinmetz M. and Meo T.: Liver cDNA probes disclose two cytochrome P450 genes duplicated in tandem with the complement C4 loci of the mouse H-2S region. *Proc. Natn. Acad. Sci. U.S.A.* **82** (1985) 4453–4457.
22. Chung B., Matteson K. J. and Miller W. L.: Cloning and characterization of the bovine gene for steroid 21-hydroxylase (P-450c21). *DNA* **4** (1985) 211–219.

23. Skow L. E., Womack J. E., Petresh J. M. and Miller W. L.: Synteny mapping of the genes for steroid 21-hydroxylase, alpha-A-crystallin, and class I bovine leukocyte(BoLA) in cattle. *DNA* **7** (1988) 143–149.
24. Higashi Y., Yoshioka H., Yamane M., Gotoh O. and Fujii-Kuriyama Y.: Complete nucleotide sequence of two steroid 21-hydroxylase genes tandemly arranged on human chromosome 6: a pseudogene and a genuine gene. *Proc. Natn. Acad. Sci. U.S.A.* **83** (1986) 2841–2845.
25. White P. C., New M. I. and Dupont B.: Structure of the human steroid 21-hydroxylase genes. *Proc. Natn. Acad. Sci. U.S.A.* **83** (1986) 5111–5115.
26. Rodrigues N. R., Dunham I., Yu C. Y., Carroll M. C., Porter R. R. and Campbell R. D.: Molecular characterization of the HLA-linked steroid 21-hydroxylase B-gene from an individual with congenital adrenal hyperplasia. *EMBO Jl* **6** (1987) 1653–1661.
27. Parker K. L., Chaplin D. D., Wong M., Seidman J. G., Smith J. A. and Schimmer B. P.: Expression of murine 21-hydroxylase in mouse adrenal glands and in transfected Y1 adrenocortical tumor cells. *Proc. Natn. Acad. Sci. U.S.A.* **82** (1985) 7860–7864.
28. Chaplin D. D., Galbreath L. G., Seidman J. G., White P. C. and Parker K. L.: Nucleotide sequence analysis if murine 21-hydroxylase genes: mutations affecting gene expression. *Proc. Natn. Acad. Sci. U.S.A.* **83** (1986) 9601–9605.
29. Chung B., Matteson K. J. and Miller W. L.: Structure of a bovine gene for P450c21 (steroid 21-hydroxylase) defines a novel cytochrome P450 gene family. *Proc. Natn. Acad. Sci. U.S.A.* **83** (1986) 4243–4247.
30. Yoshioka H., Morohashi K., Sogawa K., Yamane M., Kominami S., Takemori S., Okada Y., Omura T. and Fujii-Kuriyama Y.: 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** (1986) 4106–4109.
31. Ogata R. T. and Sepich D. S.: Murine sex-limited protein: complete cDNA sequence and comparison with murine fourth complement component. *J. Immun.* **135** (1985) 4239–4244.
32. Stavenhagen J. B. and Robins D. M.: An ancient provirus has imposed androgen regulation on the adjacent mouse sex-limited protein gene. *Cell* **55** (1988) 247–254.
33. White P. C., Chaplin D. D., Weis J. H., Dupont B., New M. I. and Seidman J. G.: The steroid 21-hydroxylase genes are located in the murine S region. *Nature* **312** (1984) 465–467.
34. Geffrotin C., Chardon P., DeAndres-Cara D. F., Feil R., Renard C. and Vaiman M.: The swine steroid 21-hydroxylase gene (CYP21): cloning and mapping within the swine leukocyte antigen locus. *Anim. Genet.* **21** (1990) 1–13.
35. Spilliaert R., Palsdottir A. and Arnason A.: Analysis of the C4 genes in baleen whales using a human cDNA probe. *Immunogenetics* **32** (1990) 73–76.
36. Levi-Strauss M., Tosi M., Steinmetz M., Klein J. and Meo T.: Multiple duplications of complement C4 gene correlate with H-2 controlled testosterone-independent expression of its sex-limited isoform, C4-Slp. *Proc. Natn. Acad. Sci. U.S.A.* **82** (1985) 1746–1750.
37. Kawaguchi H., O'hUigin C. and Klein J.: Evolution of primate C4 and CYP21 genes. In *Molecular Evolution of the Major Histocompatibility Complex* (Edited by J. Klein and D. Klein). Springer-Verlag, Heidelberg (1991).
38. Morel Y., Bristow J., Gitelman S. E. and Miller W. L.: Transcript encoded on the opposite strand of the human steroid 21-hydroxylase/complement component C4 gene locus. *Proc. Natn. Acad. Sci. U.S.A.* **86** (1989) 6582–6586.
39. Xu X. and Doolittle R.: Presence of a vertebrate fibrinogen-like sequence in an echinoderm. *Proc. Natn. Acad. Sci. U.S.A.* **87** (1990) 2097–2101.
40. Bristow J., Gitelman S. E., Shi Y. and Miller W. L.: Structure of the gene overlapping the P450c21B gene suggests it encodes a tissue-specific extracellular matrix protein. *Pediatr. Res.* **27** (1990) 76A, Abstr. No. 419.
41. Erickson H. P. and Bourdon M. A.: Tenascin: an extracellular matrix protein prominent in specialized embryonic tissues and tumors. *A. Rev. Cell Biol.* **5** (1989) 71–92.
42. Morel Y., Andre J., Uring-Lambert B., Hauptmann G., Beteuel H., Tosi M., Forest M., David M., Bertrand J. and Miller W. L.: 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** (1989b) 527–536.
43. Gitelman S. E., Bristow J. and Miller W. L.: Mechanisms and consequences of the duplication of the human C4/P450c21/Gene X locus. *Molec. Cell. Biol.* **12** (1992) 2124–2134.
44. Prentice H. L., Schneider P. M. and Strominger J. L.: C4B gene polymorphism detected in a human cosmid clone. *Immunogenetics* **23** (1986) 274–276.
45. Collier S., Sinnott P. J., Dyer P. A., Price D. A., Harris R. and Strachan T.: Pulsed field gel electrophoresis identifies a high degree of variabiity in the number of tandem 21-hydroxylase and complement C4 gene repeats in 21-hydroxylase deficiency haplotypes. *EMBO Jl* **8** (1989) 1393–1402.
46. Roth D. B. and Wilson J. H.: Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction. *Molec. Cell. Biol.* **6** (1986) 4295–4304.
47. Roth D. and Wilson J.: Illegitimate recombination in mammalian cells. In *Genetic Recombination* (Edited by R. Kucherlapati and R. Smith). American Society for Microbiology, Washington, DC (1988) pp. 621–653.
48. Konopka A. K.: Compilation of DNA strand exchange sites for non-homologous recombination in somatic cells. *Nucleic Acids Res.* **16** (1988) 1739–1758.
49. Meuth M.: Illegitimate recombination in mammalian cells. In *Mobile DNA* (Edited by D. E. Berg and M. H. Howe). American Society for Microbiology, Washington, DC (1989) pp. 833–859.
50. Henthorn P. S., Smithies O. and Mager D. L.: Molecular analysis of deletions in the human β-globin gene cluster: deletion junctions and locations of breakpoints. *Genomics* **6** (1990) 226–237.
51. Nalbantoglu J., Hartley D., Phear G., Tear G. and Meuth M.: Spontaneous deletion formation at the *aprt* locus of hamster cells: the presence of short sequence homologies and dyad symmetries at deletion termini. *EMBO Jl.* **5** (1986) 1199–1204.
52. Kozak M.: Point mutations that define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44** (1984) 283–292.
53. Hann S. R., King M. W., Bentley D. L., Anderson C. W. and Eisenman R. N.: 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** (1988) 185–195.
54. Prats H., Kaghad M., Prats A., Klagsbrun M., Lelias J., Liauzun P., Chalon P., Tauber J. P., Amalric F., Smith J. A. and Caput D.: High molecular mass forms of basic fibroblast growth factor are initiated by alternative CUG codons. *Proc. Natn. Acad. Sci. U.S.A.* **86** (1989) 1836–1840.
55. Acland P., Dixon M., Peters G. and Dickson C.: Subcellular fate of the Int-2 oncoprotein is determined by choice of initiation codon. *Nature* **343** (1990) 662–665.

56. Bernards A. and M. de la Monte S.: The ltk receptor tyrosine kinase is expressed in pre-B lymphocytes and cerebral neurons and uses a non-AUG translational inhibitor. *EMBO Jl* **9** (1990) 2279–2287.
57. Brannan C. I., Dees E. C., Ingram R. S. and Tilghlman S. M.: The product of the H19 gene may function as an RNA. *Molec. Cell. Biol.* **10** (1990) 28–36.
58. Bartolomei M. S., Zemel S. and Tilghman S. M.: Parental imprinting of the mouse. *Nature* **351** (1991) 153–155.
59. Nonaka M., Kimura H., Yeul Y. D., Yokoyama S., Nakayama K. and Takahashi M.: Identification of the 5′ flanking regulatory region responsible for the difference in transcriptional control between mouse complement C4 and Slp genes. *Proc. Natn. Acad. Sci. U.S.A.* **83** (1986) 7883–7887.
60. Polsdottir A., Arnason A., Fossdal R. and Jensson O.: Gene organization of haplotypes expressing two different C4A allotypes. *Hum. Genet.* **76** (1987) 220–224.
61. Giles C., Uring-Lambert B., Boksch W., Braun M., Goetz J., Neumann R., Mauff G. and Hauptman G.: The study of a French family with two duplicated C4A haplotypes. *Hum. Genet.* **77** (1987) 359–365.
62. Yu Y. C. and Campbell R. D.: Definitive RFLPs to distinguish between the human complement C4A/C4B isotypes and the major radgers/chido determinants: application to the study of C4 null alleles. *Immunogenetics* **25** (1987) 369–385.
63. McLean R. H., Donohue J. N., Bias W. B., Van Dop C. and Migeon C. J.: Restriction fragment analysis of duplication of the fourth component of complement (C4A). *Genomics* **2** (1988) 76–85.
64. Donohoue P. A., Jospe N., Migeon C. J. and VanDop C.: Gene conversion in salt-losing congenital adrenal hyperplasia with absent complement C4B protein. *J. Clin. Endocr. Metab.* **62** (1989) 995–1002.
65. Matteson K. J., Phillips J. A. III, Miller W. L., Chung B., Orlando P. J., Frisch H., Ferrandez A. and Burr I. M.: P450XXI (steroid 21-hydroxylase) gene deletions are not found in family studies of congenital adrenal hyperplasia. *Proc. Natn. Acad. Sci. U.S.A.* **84** (1987) 5858–5862.
66. Jospe N., Donohoue P. A., VanDop C., McLean R. H., Bias W. and Migeon C. J.: Prevalence of polymorphic 21-hydroxylase gene (CAH21B) mutations in salt losing congenital adrenal hyperplasia. *Biochem. Biophys. Res. Commun.* **142** (1987) 798–804.
67. Miller W. L.: Gene conversions, deletions, and polymorphisms in congenital adrenal hyperplasia. *Am. J. Hum. Genet.* **42** (1988b) 4–7.
68. Higashi Y., Tanae A., Inoue H. and Fujii-Kuriyama Y.: Evidence for frequent gene conversions in the steroid 21-hydroxylase (P-450c21) gene: Implications for steroid 21-hydroxylase deficiency. *Am. J. Hum. Genet.* **42** (1988) 17–25.
69. Higashi Y., Tanae A., Inoue Y., Hiroshima T. and Fujii-Kuriyama Y.: Aberrant splicing and missense mutations cause steroid 21-hydroxyase [P-450(C21)] deficiency in humans: Possible gene conversion products. *Proc. Natn. Acad. Sci. U.S.A.* **85** (1988) 7486–7490.
70. Morel Y., David M., Forest M. G., Betuel H., Hauptman G., Andre J., Bertrand J. and Miller W. L.: Gene conversions and rearrangements cause discordance between inheritance of forms of 21-hydroxylase deficiency and HLA types. *J. Clin. Endocr. Metab.* **68** (1989) 592–599.
71. Miller W. L. and Levine L. S.: Molecular and clinical advances in congenital adrenal hyperplasia. *J. Pediatr.* **111** (1987) 1–17.
72. Miller W. L.: Molecular biology of steroid hormone synthesis. *Endocrine Rev.* **9** (1988) 295–318.
73. Belt K. T., Carroll M. C. and Porter R. R.: The structural basis of the multiple forms of human complement component C4. *Cell* **36** (1984) 907–914.