Single-Nucleotide Polymorphisms in Intron 2 of CYP21P: Evidence for a Higher Rate of Mutation at CpG Dinucleotides in the Functional Steroid 21-Hydroxylase Gene and Application to Segregation Analysis in Congenital Adrenal Hyperplasia

Renée R. Jiddou, Wan-Li Wei, Kumud S. Sane, and Anthony A. Killeen

Background: Intron 2 of CYP21, the functional steroid 21-hydroxylase gene contains several single-nucleotide polymorphisms (SNPs). We tested the hypothesis that intron 2 of the pseudogene, CYP21P, might also be polymorphic and provide markers for segregation analysis of this region of the genome, including observable markers for segregation analysis of CYP21 gene deletions. A comparison of SNPs in both genes might provide insights into the rates of mutation in these duplicated genes.

Methods: After amplification with PCR, we examined restriction site polymorphisms in intron 2 of CYP21P in 24 members of the parental generation of the Centre d’Étude du Polymorphisme Humain families and selected offspring.

Results: Intron 2 of CYP21P contains frequent SNPs around nucleotide 398 and nucleotide 509, which can be typed by PCR/restriction enzyme digestion with HaeIII. Of the 48 CYP21P alleles examined, 44 could be characterized unambiguously. Of these 44 alleles, 4 were deleted, and the frequencies of restriction at the polymorphic HaeIII sites were 20 of 40 at nucleotide 398 and 30 of 40 at nucleotide 509. Both polymorphisms result from C→T transitions that occur at CpG dinucleotides. The frequencies of C at these nucleotides in CYP21P are significantly higher than at the corresponding nucleotides in CYP21 of the same individuals (P <0.01).

Conclusion: These data suggest that these CpG dinucleotides are more frequently mutated in CYP21 than in CYP21P, and that several mutations at CpG dinucleotides in the coding regions of CYP21 might result from CpG instability rather than the more usually proposed mechanism of gene conversion. These frequent SNPs provide useful markers for studying both allelic segregation of CYP21, particularly for chromosomes with known CYP21 deletions, and for investigating the origin of these polymorphisms.

The gene encoding adrenal steroid 21-hydroxylase, CYP21, is located on 6p in the MHC class III region. This region usually includes two genes, the functional CYP21 gene and a highly homologous pseudogene, CYP21P, which are arranged in tandem with the complement 4 genes (1, 2) (Fig. 1). The CYP21 and CYP21P genes overlap the tenascin-X and XA genes, and two untranslated messages termed YA and YB have also been identified as transcripts of this region (3). Mutations in CYP21 produce the most common form of congenital adrenal hyperplasia (CAH), an autosomal recessive disease with an incidence of ~1 in 13 000 newborns for the classic form, and an incidence of up to 1% in some populations for the nonclassic form (4, 5).

We have reported the existence of frequent single-nucleotide polymorphisms (SNPs) in intron 2 of CYP21, including SNPs at nucleotide (nt) 395 and nt453 (6). The former involves a C→T transition at a CpG dinucleotide. Considering the sequence homology between CYP21 and...
CYP21P, and the well-known tendency of CpG dinucleotides to undergo mutation, which leads to polymorphism among alleles (7, 8), we hypothesized that the nucleotides at the corresponding positions in CYP21P might also be polymorphic. The presence of SNPs in the pseudogene might provide markers for determining the segregation of this region of the MHC in genetic studies, especially segregation of the closely linked CYP21 in families with CAH. We have shown that CYP21 gene deletions, which account for ~20–30% of all mutations at that locus (9), can be demonstrated by an apparent failure to transmit a parental allele to an offspring as determined by SNP analysis of CYP21 in informative families (6, 10). SNPs in CYP21P might provide closely linked, observable markers for this analysis. Finally, a comparison of SNPs in CYP21P and CYP21 might provide insight into the relative rates of mutation at CpG dinucleotides in these duplicated genes.

To test these hypotheses, we examined the frequencies of SNPs which occur in CYP21P at positions homologous to those we have reported previously in CYP21 in members of the parental generation of the Centre d’Etude du Polymorphisme Humain (CEPH) families, and compared the frequencies of these SNPs in CYP21P with those at the homologous positions in CYP21.

Materials and Methods

Genotyping CYP21P in Intron 2

DNA samples from patients with CAH and their relatives were extracted from peripheral blood as described previously (11). The study was approved by institutional review boards, and informed consent was obtained from all subjects or their parents.

Intron 2 of CYP21P was amplified by 30 cycles of PCR using the following reagent concentrations: 20 mmol/L Tris-HCl, pH 8.4, 50 mmol/L KCl, 1.5 mmol/L MgCl2, 200 μmol/L each dNTP, 500 nmol/L each primer (335F and P731R; Table 1), 50–100 ng of genomic DNA, and 2.5 U of Taq polymerase (Life Technologies) in a reaction volume of 50 or 100 μL. For genomic DNAs that had been stored for several years and showed some evidence of degradation, a nested strategy was used to amplify intron 2 from CYP21P. First-round PCR reactions were performed for 30 cycles as described above, but with 500 nmol/L each of primers 1F and 822R (Table 1). For the second-round reactions, 1 μL of first-round product was used as template and the primers used were 335F and P731R (Table 1). Primer P731R is a CYP21P-specific primer that hybridizes to the region in exon 3 of CYP21P, where there is a sequence difference from CYP21 because of a deletion of 8 bp from CYP21P (1, 2). The initial denaturation of DNA was for 1 min at 94 °C. For nonnested reactions, an annealing temperature of 60 °C was used for 30 s. For nested reactions, annealing conditions were 60 °C for round 1 and 65 °C for round 2 reactions, both for 30 s. Extension conditions were 72 °C for 90 s (round 1) or 1 min (round 2).

Restriction digestion was performed by incubating 10 μL of PCR product with 5 U of HaeIII (10 U/μL; Life Technologies) in a volume of 30 μL at 37 °C for 2–3 h. The restriction solution contained the manufacturer’s 1× restriction buffer. After restriction digestion, samples were electrophoresed in 3% Metaphor agarose (FMC) containing 1× Tris-borate-EDTA buffer (0.089 mol/L Tris-borate, pH 8.3, 2 mmol/L EDTA; National Diagnostics) and 1 mg/L ethidium bromide (Sigma). Restriction products were visualized by transillumination with ultraviolet light and photographed.

Haplotype Analysis

To establish the haplotype pattern of SNPs at CYP21P in the CEPH families, reference was made to the available inheritance data for the HLA-B and HLA-DR loci in these families. These loci flank the MHC class III region (containing the 21-hydroxylase genes) on the telomeric and centromeric sides, respectively. The HLA data were used to identify second-generation siblings who have inherited different parental haplotypes in this region of 6p. On the basis of the HLA data, none of the offspring showed recombination events between the HLA-B and HLA-DR loci. Study of the transmission of parental CYP21P SNPs to these siblings enabled assignment of SNPs to individual parental haplotypes. This analysis was used to demonstrate CYP21P deletions, which occur in ~14% of individuals in the French population, to which the CEPH families belong. CYP21P deletions produce apparent failure to transmit a parental allele to an offspring when intragenic polymorphisms are studied, as we have shown previously at the CYP21 locus (6, 10).

Genotyping CYP21P of the HLA-A3, -B47, -DR7 Haplotype

The pattern of CYP21P intron 2 SNPs was studied in four unrelated individuals with the HLA-A3, -B47, -DR7 hap-

Table 1. Sequences and locations of primers used for PCR.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’ to 3’)</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>1F</td>
<td>ATG CTG CTC CTG GGC CTG C</td>
<td>Exon I</td>
</tr>
<tr>
<td>822R</td>
<td>CTC ACA GAA CTC CTG GGT CAG CTG TCT</td>
<td>Exon III</td>
</tr>
<tr>
<td>335F</td>
<td>GAA GGC ATG GTC AAA AAG TGG</td>
<td>Exon II</td>
</tr>
<tr>
<td>P731R</td>
<td>CTT TCC AGA GCA GAG ACC AAC G</td>
<td>Exon III</td>
</tr>
</tbody>
</table>
lotype, which is associated with a deletion of CYP21, and which is frequently seen in patients with classic CAH. DNA samples from parents who are heterozygous for this haplotype, and their child with CAH who is homozygous for the haplotype, were studied. In two other families, the intron 2 SNPs could be determined in parents who are heterozygous for this haplotype by segregation analysis in their offspring.

**Results**

**CYP21P INTRON 2 SNPS**

PCR amplification with primers 335F and P731R yields a product of ~400 bp. The pattern of restriction by HaeIII was studied in DNA samples from 24 members of the parental generation of the CEPH families (48 chromosomes). A diagram showing the HaeIII restriction sites in CYP21P is shown in Fig. 2. The amplified region includes three HaeIII sites in intron 2: nt398, nt509, and nt623. These correspond to nt395, nt503, and nt614 in CYP21 (2). The patterns and frequencies of restriction at these sites are shown in Table 2.

Of the 48 chromosomes, 4 were found to have deletions of CYP21P, as evidenced by the apparent failure of transmission of a parental allele to an offspring (Fig. 3). In an additional four individuals showing apparent homozygosity for markers at CYP21P, a gene deletion could not be excluded by segregation analysis because of a lack of informativeness from the family or the unavailability of offspring with the necessary combination of parental haplotypes to exclude a gene deletion. For calculations of frequencies of polymorphisms and deletions, only one allele was counted from these individuals. The CEPH study sample therefore consisted of 40 distinct CYP21P pseudogenes and 4 chromosomes with demonstrated deletions of CYP21P, a deletion rate of 4 of 44 (9.1%).

The HaeIII restriction site around nt623 was cut in all CEPH DNA samples studied, but the restriction sites at nt398 and nt509 were found to be frequently polymorphic. The nt398 site was restricted in 20 of 40 nondeletional chromosomes, and the nt509 site was restricted in 30 of 40 nondeletional chromosomes. The polymorphism between CYP21 and CYP21P at the HaeIII site around nt503/nt509 previously has been shown to result from a G→A transition on the sense strand (1, 2); however, the reason for loss of the HaeIII restriction site at nt398 in CYP21P has not been established. We confirmed by sequence analysis of three alleles that the polymorphism that destroys the latter site results from a C→T transition on the sense strand as occurs at nt395 in CYP21 (data not shown).

**CYP21P SNPS IN THE HLA-A3, -B47, -DR7 HAPLTYPE**

The pattern of HaeIII restriction in intron 2 of CYP21P in chromosomes bearing the HLA markers A3, B47, and DR7 was examined in four unrelated individuals with this haplotype. This haplotype is associated with a deletion of

---

### Table 2. Frequencies of CYP21P Intron 2 HaeIII polymorphisms and CYP21P deletions in 24 CEPH parental generation members (44 chromosomes).

<table>
<thead>
<tr>
<th>Restriction site</th>
<th>nt398 Frequency %</th>
<th>Fragments, bp</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>10 of 44 23</td>
<td>64, 109, 226</td>
</tr>
<tr>
<td>-</td>
<td>20 of 44 46</td>
<td>109, 115, 175</td>
</tr>
<tr>
<td>Deletion</td>
<td>4 of 44 9</td>
<td></td>
</tr>
</tbody>
</table>

*a* In four individuals with apparent homozygosity for polymorphisms, only one allele was counted because a gene deletion could not be excluded.

*b* + indicates restriction by HaeIII; – indicates non-restriction.

---

![Fig. 2. Restriction map of the HaeIII sites in intron 2 of CYP21P.](image)

The arrows show the locations of primers 335F and P731R in exons 2 and 3, respectively. The thicker line represents intron 2. The polymorphic HaeIII sites are indicated by *.

---

![Fig. 3. Demonstration of a CYP21P deletion in CEPH family 13292.](image)

The numbers in the first row below each subject in the pedigree diagram are CEPH sample numbers. The numbers in the second row below each subject are haplotype identifications for the HLA-B and HLA-DR loci. The pattern of restriction around nt398 and nt509 are shown below the gel. + indicates restriction, – indicates absence of restriction, and ± indicates both an allele that shows restriction and one that does not. Molecular weight markers (HaeIII digest of F+X174) are shown in lane 1. Subject 13292-01 (lane 2), who shows a pattern of restriction fragments associated with the presence of the HaeIII site at nt509 and the absence of the site at nt398, has apparently failed to transmit an allele to 13292-03 (lane 3). This is consistent with an allele bearing a CYP21P deletion. In contrast, subject 13292-02 (lane 5) has transmitted an allele that shows restriction at both sites to both offspring.
CYP21 in families with CAH. The CYP21P pseudogene on this haplotype was found to be characterized by the presence of a HaeIII restriction site at nt398 and absence of the site at nt509 in all subjects examined.

Discussion

SNPs are reported to be the most frequent form of polymorphism in humans, occurring on average every 1 kb, and their identification is one of the goals of the Human Genome Project (12, 13). The presence of SNPs in the 21-hydroxylase genes is of interest for several reasons. First, they can be used for analysis of segregation of these alleles in genetic studies, notably of families with CAH. Second, they enable identification of gene deletions in informative families, and third, they offer insights into the relative rates of mutation in these genes.

We have demonstrated previously the presence of frequent SNPs in intron 2 of CYP21, which is located ~30 kb from CYP21P, including SNPs at nt395 and nt453. The latter were demonstrated by restriction with the enzyme BstII. This enzyme did not reveal polymorphisms in CYP21P in the samples examined in this study; however, we demonstrated that there are two polymorphic HaeIII restriction sites in this intron, one located around nt398, and one located around nt509. Another HaeIII site around nt623 was restricted in all samples studied. Among non-deletional alleles, the frequency of restriction at the nt398 and nt509 sites among the CEPH samples studied was 20 of 40 alleles and 30 of 40 alleles, respectively (Table 2).

There is a significantly greater degree of polymorphism at nt398 of CYP21P than at the corresponding nt395 of CYP21 of the same subjects, as reported previously (6). Of 48 CYP21 alleles, only 8 (16.7%) were restricted by HaeIII, indicating the presence of C at nt395, whereas 20 of 40 (50%) were restricted at the homologous site among non-deletional CYP21P alleles. At nt509 in CYP21P, the HaeIII restriction site (present in 30 of 40 alleles) results from the presence of G (C on the antisense strand) at this position. In 48 CYP21 alleles from the CEPH families and 7 wild-type alleles from heterozygotes for CAH, the corresponding position (nt502) is A (T on the antisense strand) (6); thus, this polymorphism also results from a C→T transition that occurs on the antisense strand. The frequencies of these polymorphisms were significantly different between CYP21 and CYP21P at both sites ($\chi^2$, $P < 0.01$). At both positions, the C→T transitions occur at CpG dinucleotides that are situated in the sense orientation at nt395 and in the antisense orientation at nt502. CpG dinucleotides are well-recognized “hot spots” for genetic mutation involving C→T transitions (G→A on the complementary strand) (7, 8). The basis of mutation at these sites is believed to be the spontaneous deamination of 5-methylcytosine to form thymidine. On the basis of the observed higher frequency of T at nt395 in CYP21 and the apparently invariable presence of T on the antisense strand at nt502 of that gene, we suggest that the CpG dinucleotides at these positions in CYP21 are subject to a higher rate of mutation than are the corresponding nucleotides in CYP21P. Because C→T mutations at CpG dinucleotides are believed to involve methylation of cytosine, one possible explanation for the different frequencies of cytosine at these CpG dinucleotides is that the functional CYP21 is more highly methylated at these positions in the germ line than is CYP21P.

This point may have relevance to the origin of several mutations in the coding sequence of CYP21 that involve C→T transitions occurring at CpG dinucleotides. Among these are P30L (C88T), and R356W (C2108T) (9). Because these mutations have been reported in CYP21P, they frequently are considered to be evidence of gene conversion events that transfer deleterious sequences from CYP21 to CYP21P. An alternative possible etiology of these mutations, which is consistent with our observation of frequent polymorphisms at CpG dinucleotides in intron 2, is that these mutations result from an instability of CpG dinucleotides in CYP21. This mechanism may also underlie the frequent mutation at a TaqI restriction site used for distinguishing CYP21 from CYP21P on Southern blot analysis (14).

In this study, we observed a CYP21P deletion rate of 9.1%, which is in reasonable agreement with the 14% estimate of such deletions in the French population, from which the CEPH families are derived (15). A possible mechanism for deletions at either locus is the loss of genetic material arising from chromosomal misalignment during meiosis and unequal crossing over. The difference in the observed rates of gene deletion between CYP21 and CYP21P may be attributable to the selection against haplotypes bearing CYP21 deletions, which are associated with the salt-wasting form of CAH. Deletions of CYP21P appear to be silent with respect to phenotype, although a higher rate of some autoimmune diseases is associated with deletions of the neighboring C4A gene (16), which is often deleted with CYP21P.

It was of interest to examine the pattern of SNPs in the CYP21P gene in subjects having the extended MHC haplotype HLA-A3, -B47, -DR7, which is associated with a deletion of CYP21 and which frequently is present in patients with CAH. In four subjects with this haplotype, the site around nt398, but not the site around nt509, in CYP21P was restricted by HaeIII. We previously have shown that the inheritance of a deletion at CYP21 can be demonstrated by apparent failure to transmit parental CYP21 intragenic polymorphisms to an offspring (6, 10). The existence of polymorphisms at the closely linked CYP21P locus also allows for this analysis of allelic segregation in informative families, and has the advantage of providing an observable marker in both parent and offspring. The HLA-B47 antigen is known to be in strong linkage disequilibrium with a deletion of CYP21 in families with CAH. The finding of the same pattern of SNPs at the CYP21P locus in unrelated individuals with the HLA-B47 antigen is consistent with the idea that the HLA-A3, -B47, -DR7 haplotype represents a highly con-
served segment of DNA, and that the HLA-B47 antigen, which is infrequent in the general population, is possibly of relatively recent evolutionary origin (17), so that these SNPs are still conserved among apparently unrelated individuals with this haplotype.

These polymorphic restriction sites also serve as useful markers for rapid analysis of the segregation of the class III region of the MHC in family studies, and most obviously, for following the segregation of the nearby CYP21 alleles. These CYP21P markers may be especially useful for determining the segregation of haplotypes containing known CYP21 deletions, which usually requires Southern blot analysis to assess gene dosage. We previously have discussed the advantages and disadvantages of using intragenic polymorphic markers for segregation analysis of CYP21 (6, 10). In comparison with CYP21, CYP21P shows a higher frequency of deletions. Occasional haplotypes can contain duplications of one or other gene (18). These considerations must be borne in mind if these markers are used for diagnostic purposes.

References