Vol. 186, No. 1, 1992 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
July 15, 1992 Pages 256-262

THE HUMAN COMPLEMENT C4B/STEROID 21-HYDROXYLASE (CYP21) AND COMPLEMENT C4A/21-HYDROXYLASE PSEUDOGENE (CYP21P) INTERGENIC SEQUENCES: COMPARISON AND IDENTIFICATION OF POSSIBLE REGULATORY ELEMENTS[1]

Patricia A. Donohoue and Malia M. Collins

The University of Iowa, Department of Pediatrics
Division of Endocrinology

Received May 4, 1992

Summary: We determined the 1.8 kb intergenic sequences between the human complement C4B gene and the active steroid 21-hydroxylase gene in two subjects, and between the C4A gene and the steroid 21-hydroxylase pseudogene in one subject. Comparison of these sequences with each other and with published homologues revealed no differences which were unique to either intergenic region. Sequence analysis revealed two copies of an **AGGTCA** motif in all sequences. This motif is common to steroidogenic enzyme gene promoters and to the response elements for nuclear hormone receptors. Similarities with human enhancers were also found.

Introduction: Congenital adrenal hyperplasia due to 21-hydroxylase deficiency is an autosomal recessive, HLA-linked defect in adrenal steroidogenesis, which can be severe to mild in phenotype (1). Mapping and sequencing of the active 21-hydroxylase gene (CYP21) and its homologous pseudogene (CYP21P) (2,3) have provided the basis for analysis of mutations in this disorder. Tandem duplication (4,5) of complement C4A and C4B genes with the CYP21P and CYP21 genes, respectively (Figure 1), make the region likely to undergo meiotic misalignment and unequal crossingover (6). This is demonstrated in patients with

[1]Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession Numbers M94068, M94069, and M94070.

Abbreviations used: bp, base pair; C4A and C4B, A and B forms of the fourth component of the human complement pathway; CAH, congenital adrenal hyperplasia; CYP21, active steroid 21-hydroxylase gene; CYP21P, steroid 21-hydroxylase pseudogene; cDNA, complimentary DNA; DNA, deoxyribonucleic acid; HLA, human lymphocyte antigen; kb, kilobase.

0006-291X/92 $4.00

severe or salt-losing CAH by the high frequency of deletions of gene tandems, occurring in up to 30% of haplotypes (7,8,9). Such recombinations are typified by the HLA-A1,B8,DR3 haplotype, with deletion of the C4A/CYP21P tandem in CAH-unaffected homozygotes, and the HLA-A3,Bw47,DR7 haplotype, with deletion of the C4B/CYP21 tandem causing homozygotes to exhibit salt-losing CAH (10) (Figure 1). Among the known CYP21 point mutations which cause CAH, nearly all are characteristic of CYP21P, presumably having been transferred to CYP21 by a recombinational event (11). Studies have, however, failed to identify the causative mutations in up to 20% of haplotypes carrying the salt-losing CAH trait, and over 75% of haplotypes carrying milder forms of 21-hydroxylase deficiency. Some of the as yet undetected mutations may involve regulatory elements in the region upstream of CYP21.

Prior to this study, the intergenic sequence between the C4B gene and the CYP21 upstream region had not been published. In 1991, Yu (12) reported the C4A genomic sequence including the region downstream to a Bgl II site at position -1678 relative to the CYP21P initiation codon (2). We determined both the C4A/CYP21P and C4B/CYP21 intergenic sequences in order to examine for region-specific differences, and for the presence of regulatory elements which have, as yet, been uncharacterized.

Materials and Methods: The C4B/CYP21 intergenic region was isolated from two sources: 1.) a λ Charon 4A clone from a human genomic DNA library (13); and 2.) leukocyte DNA isolated as described (14) from a subject without CAH who was homozygous for HLA-A1,B8,DR3 (HLAB8). The C4A/CYP21P intergenic region was isolated from leukocyte DNA from an individual with salt-losing CAH, homozygous for HLA-A3,Bw47,DR7 (HLABw47).

Isolation and characterization of the λ clone: The λ library was first screened (15) with the C4cDNA probe pAT-F (16). Positive clones were then screened with the CYP21 cDNA probe pC21/3c (17). Bacteriophage DNA was isolated (18) from clones which hybridized with both probes. Southern analysis with the pC21/3c probe was utilized to identify intragenic restriction fragments unique to CYP21 (10). Probes were radiolabeled with [α^{32}P]dCTP using the random priming method (19). One clone, λC4B/CYP21, contained the 3.7 kb Taq I and 2.9 Kpn I fragments of the active CYP21 gene in its 13 kb insert (Figure 1).

Cloning and sequencing: The desired intergenic regions from λC4B/CYP21, HLAB8, and HLABw47 (Figure 1) underwent *in vitro* amplification with Taq Polymerase (Cetus, Norwalk, CT). The 5' primer corresponded to the 3' limit of the C4A cDNA (16), and contained a 1 base mismatch to create a Kpn I site (5'GAGTATGGCACTCAGGGGTACCAG 3'). The 3' primer included the

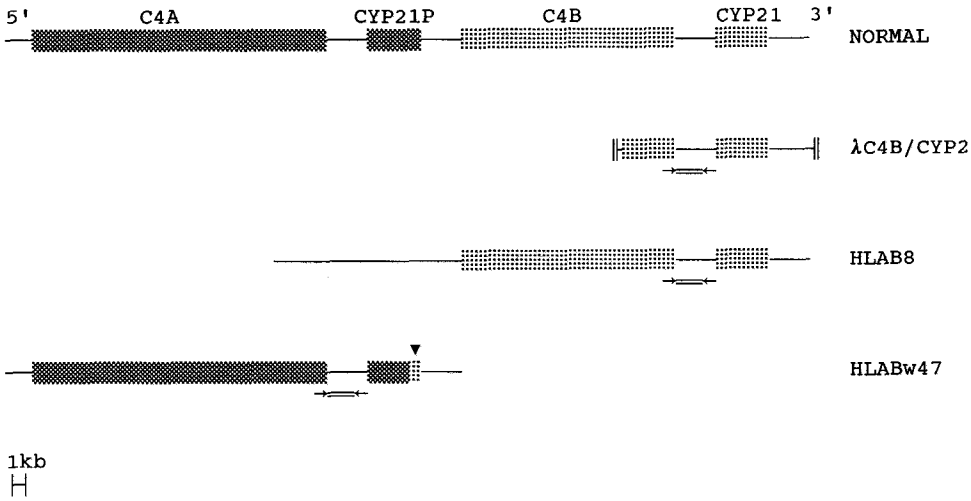


Figure 1. Gene map of the relative locations of seqences determined in this study. The top line represents the normal tandemly duplicated locus, which maps within the Class III region of the major histocompatibility complex. 5' and 3' indicate the transcriptional orientation. The second line illustrates the location of the λC4B/CYP21 insert. The third line represents the single tandem present in subject HLAB8. The fourth line illustrates the location of the single tandem present in subject HLABw47. ▼ indicates the location of sequences characteristic of CYP21. The CYP21P gene of the HLABw47 haplotype is a hybrid due to unequal crossingover in the 3' region of the gene (6). →=← indicates the location of each region which underwent *in vitro* amplification for cloning and sequencing. See text and legend for figure 2 for descriptions of HLAB8 and HLABw47.

native Kpn I site at position -1318 of CYP21 (2) (5'GATGGTACCTGAGAGGTCAGAGGC 3'). The 1.8 kb amplified products were ligated into the Kpn I site of pBSK+ (Bluescript, Stratagene, La Jolla, CA). As there was a Kpn I site within the amplified regions, the pBSK+ inserts consisted of either a 1.2 kb plus a 0.6 kb Kpn I fragment, a 1.2 kb Kpn I fragment alone, or a 0.6 kb Kpn I fragment alone. Double-stranded plasmid DNA was sequenced (20) using [α^{35}S]dATP and T7 DNA Polymerase (Sequenase 2.0 kit, U.S. Biochemical Corp., Cleveland, OH). Nested deletions for sequencing of overlapping regions were prepared by sequential digestion with Exonuclease III and Mung Bean Nuclease (Exo-Mung deletion kit, Stratagene, La Jolla, CA). In order to verify the sequence surrounding the amplified Kpn I site, the non-amplified λC4B/CYP21 DNA and the uncloned amplified DNA from HLAB8 and HLABw47 were directly sequenced (21) (TAQuence 2.0 kit, U.S. Biochemical Corp., Cleveland, OH). The sequencing primer annealed 87 bp 5' to the Kpn I site (5' GGGGCTCTGAAGACTGAG 3') and was end-labeled with [γ^{32}P]ATP and T4 polynucleotide kinase (New England Biolabs, Inc., Beverly, MA). Radionuclides were obtained from Amersham (Arlington Heights, IL); restriction enzymes and T4 DNA ligase were from Boerhinger-Mannheim (Indianapolis, IN). Sequence data were organized and analyzed using software from DNASTAR, Inc. (Madison, WI). FASTA similarity searches (22) were conducted using the GenBank On-Line service (23).

Results: The 1832 bp intergenic sequences determined in this study are shown in Figure 2. There were 16 base differences between the two C4B/CYP21 intergenic sequences determined in this study (λC4B/CYP21 and HLAB8). Of 15 base differences between λC4B/CYP21 and the C4A/CYP21P intergenic sequence determined in this study (HLABw47), 12 were common to both HLABw47 and HLAB8. The C4A published sequence differed from our C4A/CYP21P (HLABw47) sequence by 18 bases. Within the region 3' to the Bgl II site at base 1470, this site being the 5' limit of the published CYP21 sequence (2), λC4B/CYP21 differed from CYP21 by 3 bases, and from CYP21P by 2 bases. The latter two base differences were common to both CYP21 and CYP21P. No sequences unique to either intergenic region were identified.

Sequence analysis revealed two copies of an exact AGGTCA motif at bases 866 and 1566. Comparison of the λC4B/CYP21 sequence with sequences of human enhancers identified four loci which shared >50% but <60% similarity: the polyoma enhancer fragment A (24), the epidermal growth factor receptor proto-oncogene downstream enhancer and the epidermal growth factor upstream enhancer (25), and the plasmid stimulating X-chromosome sequence (26).

Discussion: The intergenic sequences determined in this study will be useful in the analysis of control of CYP21 gene expression, and of mutations which produce 21-hydroxylase deficiency. The AGGTCA motif identified in the intergenic regions described above is shared by the promoter elements of murine side-chain cleavage enzyme (P-450SCC), 11ß-hydroxylase (P-45011ß), and 21-hydroxylase (P-450C21) (27). This motif is also the response element for several nuclear hormone receptors. Analysis of the published CYP21 genomic sequence in the region downstream of that determined in our study revealed a third AGGTCA motif at base -241 relative to the transcription initiation site (2). The AGGTCA motif and related sequences are required for normal expression of the above murine steroidogenic enzyme genes, and the same nuclear protein is likely to interact with this motif in the promoter region of all three genes (27). A related motif (CCAAGGTC) has been identified as a cAMP response element in the bovine 11ß-hydroxylase gene (28). These authors identified the same motif in the sequences of the 5' flanking regions of bovine, murine and human P-450SCC, 11ß, and C21. The significance of multiple copies of the AGGTCA motif in the human

```
CCAGGTGTGA GGGCTGCCCT CCCACCTCCG CTGGGAGGAA CCTGAACCTG GGAACCATGA 60  λC4B/CYP21

AGCTGGAAGC ACTGCTGTGT CCGCTTTCAT GAACACAGCC TGGGACCAGG GCATATTAAA 120

GGCTTTTGGC AGCAAAGTGT CAGTGTTGGC AGTGAAGTGT CAGTGTGTGT TGCTAGGGCT 180
                                  C                                  HLAB8
                                  C                                  HLABw47
                                  C                                  C4A

GAGAGCAGTG CCCCTGCCCG ATGCAGTTCT GGGCAGGCCA GGTTGACATA ACCTTAGACT 240

CTCTGAGCCC TGATGACCCT TGGCTGTTCA GCTCTGCTAG AACCTCCCAG ATGACCCGCT 300
                      G↑                                             C4A

AGAGTCTAGT GCTTCACAGG ACCACCCCGA GCAGAACTGG GACCCAAGAG CCTGCACCCC 360
 G↑                                                                  C4A

AAGGACCAGA GTCATGCCAA GACCACCCTT CAGCTTCCAA GGCCCTCCAC TGCCCGGCTG 420
            C↑                                          T --        HLAB8
            C↑                                          T --        HLABw47
            C↑                                                      C4A

TCGCCAGTCA CCACGGCCTC AGACAGGGCT TGTGCTCAGC TGACACCTGT GACACAGCTC 480

TTCTGCCTCA TGAGCTGTTG TCCAGCTACA CCTCCCCGAC TCTGTCCTCG TGCTGCTGGC 540

GGTTCTGAGG TCTGCAGATT TTAGCTGAGT TCCGGGCTGT TGAAAGCCTG CTGACGCTTG 600
                                                C↑   --             HLAB8
                                                C↑   --             HLABw47

GTTCTGTTAT CAGTGGAATG AGGTGACTTT CCCGGAGTTG TGCAATCCTC AGGTCCGGCA 660
              A                                                     HLABw47
                                      C↑                            C4A

GTGTCTTCTT CCAGTTACTG GTTTCAAACA AGCCAAAAGT CTGACTTTGG TGTGTTTGTG 720
 X                                                                  HLABw47
```

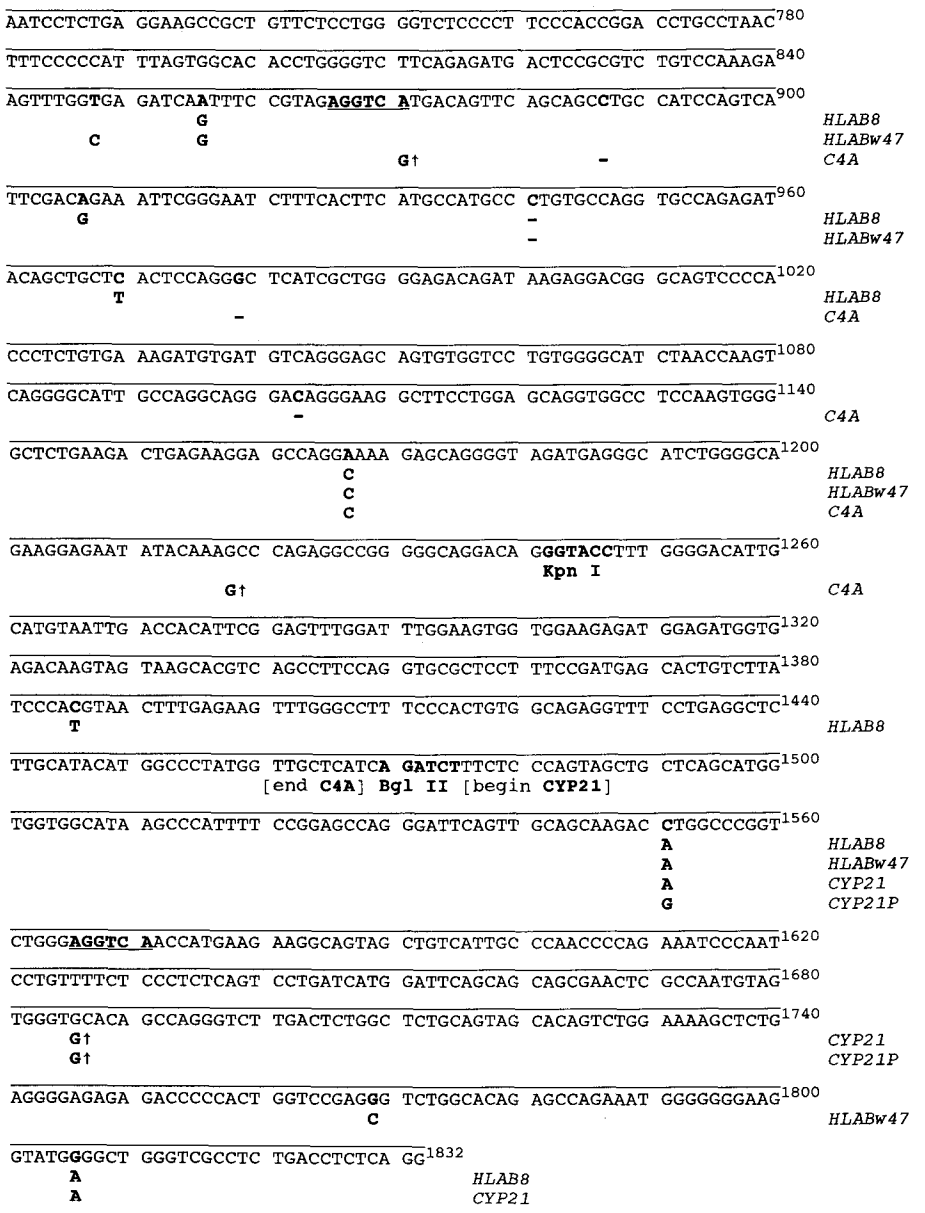
Figure 2. C4B/CYP21 and C4A/CYP21P intergenic sequences. The 1832 base sequence of a bacteriophage λ clone containing the C4B/CYP21 intergenic region (λC4B/CYP21) is shown in its entirety. The termination codon for C4 is underlined (bases 8-10). Differences between λC4B/CYP21 and that of the homologous region from other individuals are indicated below the sequence of λC4B/CYP21. Single base differences are indicated by the appropriate abbreviation directly below the corresponding position (indicated by boldface type) in the λ sequence. Base insertions are indicated by the base abbreviation followed by ↑. The base is inserted immediately prior to the base above the ↑. Base deletions are indicated by - beneath the site of the deletion. The sequences containing the base differences shown are indicated at the right margin on the line containing the indicated difference. The intergenic Kpn I site is shown in boldface type at base 1242, and the Bgl II site which divides the published genomic sequences of C4A and CYP21 is shown in boldface type at position 1470. The hexanucleotide motif identical to that of the promoter element shared by three murine steroidogenic enzymes (27) is shown at bases 866 and 1566, underlined and boldfaced. The **X** below base 662 indicates a base which could not be read in the HLABw47 sequence. Sequences with which λC4B/CYP21 was compared include two sequences determined in this study and two published sequences:

HLAB8: PCR-amplified DNA corresponding to the C4B/CYP21 intergenic region from a subject without 21-hydroxylase deficiency, who was homozygous for HLA-A1, B8, DR3 and deletion of the C4A/CYP21P tandem.

HLABw47: PCR-amplified DNA corresponding to the C4A/CYP21P intergenic region from a subject with salt-losing 21-hydroxylase deficiency who was homozygous for HLA-A3, Bw47, DR7 and deletion of the C4B/CYP21B tandem.

C4A: The C4A/CYP21P intergenic sequence which extends to the BglII site 5' to the CYP21P coding region (12).

CYP21 and CYP21P: The sequences upstream of the CYP21 and CYP21P coding regions, respectively, beginning with the BglII site at base 1470 (2).



CYP21 upstream region is unknown, however future studies to determine their relative importance will now be possible. It is intriguing that one or more of these motifs may be involved in the coordinated regulation of multiple steroidogenic enzyme gene expressions. The significance of similarity with human enhancers is unknown.

Acknowledgments: P.A.D. is a Carver Clinician Scientist of the University of Iowa. This work was supported by the Roy J. Carver

Charitable Trust and NIH DK41260, 00180, and 25295. The contributions of Doctors Cornelis Van Dop, Claude Migeon, Craig Porter, and Jeff Murray were greatly appreciated.

References:

1. Migeon, C.J., and Donohoue, P.A. (1991) Endocrinology and Metabolism Clinics of North America 20, 277-296.
2. Higashi, Y., Yoshioka, H., Yamane, M., *et al.* (1986) Proc. Natl. Acad. Sci. U.S.A. 83, 2841-2845.
3. White, P.C., New, M.I., and Dupont, B. (1986) Proc. Natl. Acad. Sci. U.S.A. 83, 5111-5115.
4. Carroll, M.C., Campbell, R.D., and Porter, R.R. (1985) Proc. Natl. Acad. Sci. U.S.A. 82, 521-525.
5. White, P.C., Grossberger, D., Onufer, B.J., *et al.* (1985) Proc. Natl. Acad. Sci. U.S.A. 82, 1089-1093.
6. Donohoue, P.A., Jospe, N., Migeon, C.J., and Van Dop, C. (1989) Genomics 5, 397-406.
7. Jospe, N., Donohoue, P.A., Van Dop, C., *et al* (1987) Biochem. Biophys. Res. Commun. 142, 798-804.
8. White, P.C., Vitek, A., Dupont, B., and New, M.I. (1988) Proc. Natl. Acad. Sci. U.S.A. 85, 4436-4440.
9. Morel, Y., Andre, J., Uring-Lambert, B., *et al.* (1989) J. Clin. Invest. 83, 527-536.
10. Donohoue, P.A., Jospe, J., Migeon, C.J., *et al.* (1986) Biochem. Biophys. Res. Commun. 136, 722-729.
11. Higashi, Y., Tanae, A., Inoue, H., *et al.* (1988) Proc. Natl. Acad. Sci. U.S.A. 85, 7486-7490.
12. Yu, C.Y. (1991) J. Immunol. 146, 1057-1066.
13. Maniatis, T., Hardison, R.C., Lacy, E., *et al.* (1978) Cell 15, 687-701.
14. Donohoue, P.A., Van Dop, C., McLean, R.H., *et al.* (1986) J. Clin. Endocrinol. Metab. 62, 995-1007.
15. Benton, W.D., and Davis, R.W. (1977) Science 196, 180-182.
16. Belt, K.T., Carroll, M.C., and Porter, R.R. (1984) Cell 36, 907-914.
17. White, P.C., New, M.I., and Dupont, B. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 7505-7509.
18. Davis, L.G., Dibner, M.D., and Battey, J.F. (1986) *Basic Methods in Molecular Biology*, Elsevier, New York, pp. 158-165.
19. Feinberg, A.P., and Vogelstein, B. (1984) Anal. Biochem. 137, 266-267.
20. Kraft, R., Tardiff, J., Krauter, K.S., and Leinwand, L.A. (1988) BioTechniques 6, 544-547.
21. Johnstone, E.M., Smith, D.P., Hsuing, H.M., and Little, S.P. (1991) Comments (U.S. Biochemical Corp.) 18, 22-23.
22. Pearson, W.R., and Lipman, D.J. (1988) Proc. Natl. Acad. Sci. U.S.A. 85, 2444-2448.
23. Bilofsky, H.S., and Burks, C. (1988) Nucl. Acids Res. 16, 1861-1864.
24. Weber, F., deVilliers, J., and Shaffner, W. (1984) Cell 36, 983-992.
25. Maekawa, T., Imamoto, F., Merlino, G.T., *et al.* (1989) J. Biol. Chem. 264, 5488-5494.
26. Riley, D.E., Reeves, R., and Gartler, S.M. (1986) Nucleic Acids Res. 14, 9407-9423.
27. Rice, D.A., Mouw, A.R., Bogerd, A.M., and Parker, K.L.: Molec. Endocrinol. 5, 1552-1561.
28. Honda, S-I., Morohashi, K-I., Omura, T. (1990) J. Biochem. (Tokyo) 108, 1042-1049.