Deficiencies of Human Complement Component C4A and C4B and Heterozygosity in Length Variants of RP-C4-CYP21-TNX (RCCX) Modules in Caucasians: The Load of RCCX Genetic Diversity on Major Histocompatibility Complex-associated Disease

By Carol A. Blanchong,*‡ Bi Zhou,* Kristi L. Rupert,*§ Erwin K. Chung,* Karla N. Jones,‡ Juan F. Sotos,‡ William B. Zipf,‡ Robert M. Rennebohm,‡ and C. Yung Yu*‡§

From the *Children’s Research Institute, Columbus, Ohio 43205; and the ‡Department of Pediatrics, §The Ohio State Biochemistry Program, the ‡Department of Molecular Virology, Immunology and Medical Genetics, and the †Comprehensive Cancer Center, College of Medicine and Public Health, The Ohio State University, Columbus, Ohio 43201

Abstract

The complement component C4 genes located in the major histocompatibility complex (MHC) class III region exhibit an unusually complex pattern of variations in gene number, gene size, and nucleotide polymorphism. Duplication or deletion of a C4 gene always conurs with its neighboring genes serine/threonine nuclear protein kinase RP, steroid 21-hydroxylase (CYP21), and tenascin (TNX), which together form a genetic unit termed the RCCX module. A detailed molecular genetic analysis of C4A and C4B and RCCX modular arrangements was correlated with immunochemical studies of C4A and C4B protein polymorphism in 150 normal Caucasians. The results show that bimodular RCCX has a frequency of 69%, whereas monomodular and trimodular RCCX structures account for 17.0 and 14.0%, respectively. Three quarters of C4 genes harbor the endogenous retrovirus HERV-K(C4). Partial deficiencies of C4A and C4B, primarily due to gene deletions and homoexpression of C4A proteins, have a combined frequency of 31.6%. This is probably the most common variation of gene dosage and gene size in human genomes. The seven RCCX physical variants create a great repertoire of haplotypes and diploid combinations, and a heterozygosity frequency of 69.4%. This phenomenon promotes the exchange of genetic information among RCCX constituents that is important in homogenizing the structural and functional diversities of C4A and C4B proteins. However, such length variants may cause unequal, interchromosomal crossovers leading to MHC-associated diseases. An analyses of the RCCX structures in 22 salt-losing, congenital adrenal hyperplasia patients revealed a significant increase in the monomodular structure with a long C4 gene linked to the pseudogene CYP21A, and bimodular structures with two CYP21A, which are likely generated by recombinations between heterozygous RCCX length variants.

Key words: complement C4 deficiency • congenital adrenal hyperplasia • endogenous retrovirus • gene dosage • heterozygosity

Introduction

A distinct feature of the immune system is the drive to achieve and maintain great diversities. Immunoglobulins accomplish this goal at the local level in B lymphocytes by somatic recombination of gene segments and hypermutations. The MHC class I and class II molecules achieve their diversities at the germline level, resulting in an enormous collection of allelic, polymorphic variants within a population (1). The variety of the human complement C4 proteins is generated through a complex pattern of genetic differences in gene size, gene number, and nucleotide polymorphisms (for reviews, see references 2, 3).
Complement component C4 is an essential component for the effector arm of the humoral immune response. The two isotypes C4A and C4B have >41 variants detectable by gross differences in electrophoretic mobility (4). C4A and C4B demonstrate differential chemical reactivities, as C4A displays higher affinity for amino group-containing antigens or immune complexes, and C4B for hydroxyl group-containing antigens (5–8). C4A is generally associated with the Rodgers (Rg) blood group antigens and C4B, the Chido (Ch) blood group antigens (9–11). The isotype specificity of C4A and C4B proteins is defined by four amino acid substitutions at residues 1101–1106, by which C4A has PCVLD and C4B has LSPV1H (11). The Rg1 and Ch1 are the results of two amino acid substitutions at residues 1188–1191. The Rg1 epitope has VDL.L, and the Ch1 epitope has ADLR (12), which can be recognized by mAbs (13). The nucleotide substitutions leading to these changes are recognizable by restriction fragment length polymorphisms (RFLPs) in Southern blot analyses (10).

In a haploid genome, there are generally two C4 genes in tandem coding for C4A and C4B, but deletions or duplications of C4 genes are well documented (14–16). The deletion or duplication of a C4 gene is always accompanied by its three neighboring genes: serine/threonine nuclear protein kinase R, peroxisome proliferator activated receptor gamma, and alpha-1 antitrypsin (14). The frequency of a partial C4A or C4B protein deficiency in the normal Caucasian population was previously estimated to be between 25.5 and 33.5% (23). The genetic etiology may be due to the deletion of a specific C4A or C4B gene (14), the presence of two genes coding for two identical C4 isotypes or allotypes (10, 24, 25), or the presence of pseudogenes caused by point mutations (26, 27). Partial C4 deficiency is the most common immune protein deficiency in humans. The varying efficiencies of complement activation that results from partial or complete deficiency may have a profound effect on the susceptibility of an individual to autoimmune diseases. An above average efficiency may increase the risk of tissue damage in autoimmunity, and a below average efficiency may slow the dissolution and removal of immune aggregates through erythrocyte complement receptor CR1 (28). The frequency for partial C4A deficiency is significantly increased in patients with the autoimmune disease SLE from different ethnic groups (29, 30; for a review, see reference 31). Although the bases for C4 deficiencies in specific individuals have been elucidated, the causes for the high frequencies of C4A and C4B null alleles in the normal and in the disease populations have not been systematically analyzed. An accurate account of the epidemiology of C4A and C4B deficiencies and polymorphisms is important for understanding the roles of complement C4 on autoimmunity. Similarly, elucidation of the RCCX modular variations in the population may cast light on the mechanism of pathogenesis of the disease congenital adrenal hyperplasia (CAH), which manifests a high range of disease severities from the mild simple virilizing phenomenon to the life-threatening salt-losing defects (32).

The presence and the dosage of C4A and C4B genes together with three other constituent genes of the RCCX modules in 150 Caucasians were investigated. The phenotypes of the C4 proteins were analyzed by immunochromatographic experiments. From these analyses, we have uncovered a high frequency of trimodular RCCX modules (i.e., three C4 genes in a haplotype), elucidated the molecular bases of C4A and C4B partial deficiencies, and observed a very high frequency of heterozygosity in length variants of the RCCX modules. Application of the similar study on patients with CAH revealed specific patterns of RCCX modules in the patient population.

Materials and Methods

Study Populations. The normal group comprised 150 healthy Caucasian females recruited from central Ohio. None of the volunteers had a personal history or a first to second degree relative family history of autoimmune disease. The CAH group comprised 22 salt-losing CAH patients recruited from the Endocrinology Clinics of the Columbus Children’s Hospital, Columbus, OH. Informed consent from each participant was obtained according to approved protocol by the Institutional Human Subject Review Board of Columbus Children’s Hospital. From each individual, 10 ml of peripheral blood was taken and EDTA was added to a final concentration of 1 mM.

Isolation of G enomic DNA and Southern Blot Analysis. Genomic DNA was isolated from the blood samples using the Puregene DNA isolation kit (Genta Systems, Inc.). 8 μg of genomic DNA was digested to completion with appropriate restriction enzymes. Restriction enzymes used included TaqI (GIBCO BRL), BamHI, PstI, EcoO109I, and NlaIV (New England Biolabs). DNA fragments were resolved by electrophoresis of 0.8% agarose gels except for EcoO109I- and NlaIV-digested DNA, for which 1.2% gels were used. Southern blot analysis was performed as described previously (10, 17).

Abbreviations used in this paper: CAH, congenital adrenal hyperplasia; Ch, Chido; CYP21, steroid 21-hydroxylase; L, long, monomodular RCCX with single long C4 gene; LL, bimodular RCCX with two long C4 genes; LLL, trimodular RCCX with three long C4 genes; LS, bimodular RCCX with one long and one short C4 gene; LSS, trimodular RCCX with one long and two short C4 genes; RCCX, RP-C4-CYP21-TNX; RFLP, restriction fragment length polymorphism; Rg, Rodgers; RP, serine/threonine nuclear protein kinase; S, short, monomodular RCCX with single short C4 gene; TNX, tenascin.
Results

An intensive study of the molecular organization of the RCCX modular structure and the genotype and phenotype of complement component C4 was performed. Definitive RFLPs were applied to detect and distinguish the number of RCCX modules present on each chromosome 6, the length polymorphism of the C4 genes, the identity of C4A and C4B isotypes, and the association with major antigenic determinants of the Rg1 and Ch1 blood group. The phenotypes of complement component C4 were examined by immunofixation of EDTA-blood plasma. The antigenic determinants of the C4A and C4B proteins were determined by immunoblot analysis using anti-Rg1 and anti-Ch1 mAbs. The genotypic and phenotypic data were then analyzed together to interpret the RCCX and the probable C4 haplotypes of each individual.

Patterns of RCCX Variations in Selected Individuals

12 samples were selected from the study population to exemplify the complexity of the RCCX modular variations and the corresponding genotypes and phenotypes of C4 (Fig. 2). A descriptive summary of these subjects is provided in Table I. In brief, these samples can be categorized into five groups.

Bimodular RCCX Haplotypes with Both C4A and C4B Proteins

The genotype and phenotype of C006 are described in detail to illustrate the complex analysis of the data. In Fig. 2, panel I-A, the Taq1 Southern blot demonstrates equal band intensities for the 7.0-kb fragment corresponding to RP1-C4L and the 6.0-kb fragment corresponding to RP2-C4L. In addition, the relative band intensities for the 3.7-kb fragment for CYP21B and the 3.2-kb fragment for CYP21A, as well as the 2.5-kb fragment for TNXB and 2.4-kb fragment for TNXA, are equal. These data suggest a homozygous, bimodular haplotype of C4A-C4L-CYP21A-TNXA-RP1-C4L-CYP21B-TNXB, abbreviated LL/LL. Confirmation of the bimodular structure is provided by the BamHI RFLP (panel I-B) demonstrating equal intensities of the 6.5-kb band for TNXB and the 4.9-kb band for BamHI band for TNXA.

In panel II-A, the PshAI Southern blot differentiates isotype C4A (4.35-kb fragment) from C4B (8.2-kb fragment). The respective bands in C006 are of similar intensities, reflecting the presence of both isotypes to the same degree. EcoO109I (panel II-B) defines the genotype based on association with C4-Rg1 (565-bp fragment) and C4-Ch1 (458-bp fragment). The C4 allotyping and immunoblot analyses reveal that C006 has equal protein expression of C4A3 and C4B1 (panel III-A), which are associated with Rg1 and Ch1 epitopes (panel III-B), respectively. The bimodular LL/LL structure suggests that C006 has two C4 genes on each chromosome 6, and the possible C4 haplotypes are C4A3 B1/C4A3 B1 or C4A3 A3/C4B1 B1.

Figure 1. A schematic diagram of the human MHC complement gene cluster. The functional genes of the RCCX modules are highlighted. A horizontal arrow represents the transcriptional orientation of a gene; vertical arrows represent locations of the DNA hybridized to probes A–F. K(C4), endogenous retrovirus HERV-K(C4); 21A, CYP21A; 21B, CYP21B; XA, TNXA.
Based on the presumption that within a bimodular RCCX structure with equal quantities of C4A and C4B genes and proteins, the first C4 locus codes for C4A and the second C4 locus codes for C4B, C006 is assigned a probable haplotype of C4A3 B1/C4A3 B1.

Patient C019 is assigned bimodular, LL/LS. This assignment is based on the presence of a 7.0-kb TaqI fragment (RP1-C4L), a 6.0-kb fragment (RP2-C4L), and a 5.4-kb fragment (RP2-C4S), with a band intensity for the 7.0 kb fragment that is two times that of the 6.0- and 5.4-kb fragments (panel I-A, lane 2). Band intensities for TNXA and B and for CYP21A and B on the TaqI digest and for TNXA and B on the BamHI digest are similar, confirming a bimodular/bimodular structure as in C006. C019 has C4 allotypes A4, A6, B1, and B1 (lane 2) and her haplotypes are assigned C4A4 B1/C4A6 B1. The C4A and C4B proteins have the expected association with the Rg1 epitope and the Ch1 epitope, respectively.

For patient C144, TaqI and BamHI RFLPs suggest that she has a homozygous LS/LS RCCX modular structure. The PshAI Southern blot demonstrates an equal number of C4A and C4B genes (panel II-A, lane 3). However, the EcoO109I blot for Rg1 and Ch1 of the C4 genes reveals a ratio of 1:3 (panel II-B, lane 3), suggesting that one of the C4A proteins is associated with Ch1. The C4 allotyping data confirmed three C4 proteins reacting with the Ch1 mAb: C4A12, B1, and B2. The C4 allotype reacting with Rg1 mAb is C4A4. The predictive haplotypes for C144 are C4A4 B1/C4A4 B2 or C4A12 B2/C4A4 B1.

These four individuals (C006, C088, C019, and C144) each have a bimodular/bimodular RCCX structure containing four C4 genes in each diploid genome. The C4 gene in the first locus is always long. The C4 gene in the second locus is either long or short. These C4 genes encode for both isotypes of C4 protein, C4A and C4B.

Individuals with Complete or Partial C4A Protein Deficiency (C4AQ0 Phenotypes). The next group of individuals (Fig. 2, lanes 4–6) all demonstrate complete (C071 and C040) or partial (C098) deficiency of C4A protein, verified on the C4 allotyping and immunoblot experiments (panels III-A and III-B). Each of these individuals demonstrates the presence of a 6.4-kb fragment on the TaqI RFLP, corresponding to a monomodular RCCX structure with a single short C4B gene linked to RP1, or RP1-C4B-CYP21B-TNXB. C071 has this haplotype on both of her chromosomes, and absence of C4A, CYP21A, TNXA, and RP2 (lane 4). Her genotype data based on PshAI and EcoO109I RFLPs demonstrate the absence of C4A genes and C4 genes associated with Rg1 (panels II-A and II-B). The allotype gel confirms the absence of C4A protein expression (panel III-A). RFLP analysis of C040 and C098 with TaqI reveals a bimodular RCCX structure with 7.0- and 6.0-kb fragments consistent with the bimodular RP1-C4L-CYP21A-TNXA-RP2-C4S-CYP21B-TNXB (abbreviated LS) in addition to the
6.4-kb fragment mentioned above. C4 genotyping in panels II-A and II-B demonstrates the presence of C4A and C4B genes and their associations with Rg1 and Ch1, respectively. However, the allotyping experiment shows no C4A protein detectable in C040 (panel III-A, lane 5). Sequence-specific PCR revealed that C040 has a C4A mutant gene characterized by a 2-bp insertion in codon 1213 of exon 29 (data not shown).

Therefore, the deficiency of C4A proteins, as demonstrated in these three individuals, may be caused by the presence of a monomodular RCCX structure with the absence of a C4A gene (gene deletion), or by the presence of a C4A pseudogene (as in C040).

Individuals with Complete or Partial C4B Protein Deficiency (C4BQ0 Phenotypes). C4 allotyping and immunoblot experiments of the next group consist of three individuals (lanes 7–9) who demonstrate a complete (C020) or a partial (C106 and C015) deficiency of C4B. The modular structures of C020, as shown by TaqI and BamHI RFLPs (panel I, lane 7), are bimodular LL/bimodular LL, which imply the presence of four C4 genes. However, no C4B genes are detectable on PshAI RFLP and no Ch1-associated C4 genes are present in the EcoO109I blot. Allotyping experiments suggested that all four C4 genes in C020 result in production of C4A proteins A3, A2, A2, and A5 (panel III, lane 7). In other words, the complete C4B deficiency in C020 is due to all four C4 gene loci from the bimodular/bimodular RCCX haplotypes encoding for C4A proteins. TaqI and BamHI RFLPs reveal that in C106, the restriction fragments for RP1-C4(L), CYP21B, and TNXB are two times more intense than those for RP2-C4(L), CYP21A, and TNXA (panels I-A and I-B, lane 8). Therefore, C106 has the heterozygous bimodular LL and monomodular L haplotypes. The PshAI RFLP reveals that there are two C4A genes and one C4B gene. The EcoO109I RFLP shows that there are two C4 genes associated with Rg1 and one C4 gene associated with Ch1. Allotyping and immunoblot experiments reveal a corresponding 2:1 ratio for C4A3-Rg1 and C4B1-Ch1. The partial C4B deficiency is due to the deletion of a C4B gene associated with a monomodular structure.

The TaqI and BamHI RFLPs revealed that C015 has the LL/LS RCCX structures. However, PshAI and EcoO109I RFLPs revealed three C4A or Rg1-associated genes, and one C4B or Ch1-associated gene (panel II, lane 9). C4 allotyping and immunoblot experiment revealed a 3:1 ratio for C4A3-Rg1 and C4B6-Ch1. Therefore, C015 is assigned C4A3 A3/C4A3 B6. The apparent partial C4B de-
ficiency is due to the expression of two C4A3 allotypes from one of the chromosome 6.

Trimodular RCCX Structures. Phenotypic studies of C4 on C007 and C009 show a significant higher expression of C4A3-Rg1 than C4B1-Ch1 (Fig. 2, panel III, lanes 10 and 11). However, RFLP studies for the RCCX and C4 genotypes do not suggest any monomodular RCCX structures in these individuals.

On the TaqI RFLP, C007 demonstrates a 7.0-kb fragment corresponding to RP1-C4L, a 6.0-kb fragment corresponding to RP2-C4L, and a 5.4-kb fragment corresponding to RP2-C4S. These three TaqI fragments appear to have a ratio of 2:2:1. Differential gene dosages of CYP21 and TNX are demonstrated, as the 3.2-kb fragment for CYP21A is more intense than the 3.7-kb CYP21B, and the 2.4-kb TNX is greater than 2.5-kb TNXB. In addition, the 5.0-kb BamHI fragment for TNX and the 6.5-kb fragment for TNXB have a ratio of ~3:2 (panel I-B, lane 10). Genotyping of C4 revealed that C007 has more C4A and Rg1-associated genes than C4B or Ch1-associated genes (Fig. 2, panels II-A and II-B, lane 10). As there are two RP1 and two TNXB genes in every individual, the higher gene dosages of RP2 and TNX suggest that one of the chromosome 6 has a trimodular RCCX structure with three long C4 genes (Fig. 2, panel III, lanes 10). The presumed RCCX haplotypes for C007 are trimodular RP1-C4A3(L)-CYP21A-TNXA-RP2-C4A3(L)-CYP21A-TNXA-RP2-C4B1(S)-CYP21B-TNXB and bimodular RP1-C4A3(L)-CYP21A-TNXA-RP2-C4B1(S)-CYP21B-TNXB.

Results of the RFLP studies for C007 reveal homzygous trimodular structures LLL/LLL. The 6.0-kb RP2-C4L fragments are two times more intense than the 7.0-kb RP1-C4L fragments. In addition, the band intensities for fragments corresponding to CYP21A and TNXA both have 2:1 ratios over their corresponding CYP21B and TNXB. Since the C4A or Rg1-associated genes are higher in dosages than the C4B or Ch1-associated genes, the C4 haplotypes of C007 are assigned C4A3 A3 B1/C4A3 A3 B1.

Unusual RCCX Patterns. C046 demonstrates two C4 genes by the presence of the 7.0- and 6.4-kb TaqI fragments, indicating a monomodular L and a monomodular S structure, respectively (panel I, lane 12). The TaqI and BamHI RFLPs show the absence of TNXA-RP2 gene segments but the presence of a CYP21A and a CYP21B. On one chromosome there is a deletion of TNXA-RP2-C4B1-CYP21B; on the other chromosome there is a deletion of C4A-CYP21A-TNXA-RP2. This individual is a carrier of CAH because of the CYP21B gene deletion. Another interesting point is the PshAI RFLP reveals that both C4A and C4B genes are present (panel II-A, lane 12), but on the EcoO109I blot, both C4 genes are associated with Ch1 (panel II-B, lane 12). The allopying gel and immunoblot experiments confirmed the presence of two C4 proteins expressing the Ch1 epitope. The predicted C4 haplotypes are A00 B1/A1 B00. Similar to the C4A12 in C0144, C4A1 has the reverse association with Ch1. The C4A00 or C4B00 in each case is due to the deletion of an RCCX module.

Table II. Genotypes and Phenotypes of Complement C4A and C4B in the Normal Caucasian Population

<table>
<thead>
<tr>
<th>Phenotypes of C4A and C4B</th>
<th>C4 genotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allotype</td>
<td>n</td>
</tr>
<tr>
<td>A1</td>
<td>1</td>
</tr>
<tr>
<td>A12</td>
<td>1</td>
</tr>
<tr>
<td>A2</td>
<td>17</td>
</tr>
<tr>
<td>A3</td>
<td>274</td>
</tr>
<tr>
<td>A4</td>
<td>11</td>
</tr>
<tr>
<td>A5</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>322</td>
</tr>
</tbody>
</table>

*Frequency among all C4 allotypes.
by the concurrent presence of a 4.9-kb TNXA fragment and a 6.5-kb TNXB fragment in the BamHI RFLP. Further supporting data include but are not limited to the equal intensities of the restriction fragments for CYP21A and CYP21B, and/or for C4A and C4B, C4-Rg1 and C4-Ch1. Among the 300 copies of chromosome 6 (or haplotypes) investigated, there are 138 (46.0%) chromosomes with a bimodular LL haplotype and 69 (23%) chromosomes with a bimodular LS haplotype. Altogether, 69.0% of the chromosome 6 are bimodular in RCCX structures.

The presence of a monomodular RCCX structure with a single short C4 gene is demonstrated by the presence of a 6.4-kb fragment for RP1-C4(S) accompanied by a copy of CYP21B (3.7 kb) and a copy of TNXB (2.5 kb). The presence of a monomodular RCCX structure with a long C4 gene is demonstrated by the presence of the 7.0-kb RP1-C4(L) and the absence of the 6.0-kb RP2-C4(L) or the 5.4-kb RP2-C4(S) in a TaqI RFLP, and the absence of 4.9-kb TNXA and RP2 in a BamHI RFLP. In the study population, there are 33 (11%) monomodular short chromosomes and 18 (6%) monomodular long chromosomes. Together, the two monomodular RCCX haplotypes have a frequency of 17%.

In contrast to the monomodular RCCX haplotypes, the trimodular structure is defined by relatively higher intensities of the restriction fragments for RP2-C4(L) and/or RP-C4(S) compared with that for the RP1-C4(L) in a TaqI RFLP. Similarly, the ratio of RP1 to RP2 and of TNXB to TNXA is <1:1 in a BamHI RFLP. In the study population, 19 (6.33%) chromosomes are LSS, 22 (7.33%) chromosomes are LLL, and 1 (0.33%) is LLS (or LSL). As shown in Fig. 3, seven RCCX length variants, LL, LS, L, S, LLL, LSS, and LLS, are detectable. These 7 modular structures give rise to 20 different haplotypes of RP, C4, CYP21, and TNX genes.

**RCCX MODULES**

**Bimodular 69.0%**

- LL 46.0%
  - 1. CA4 31A COB 21B
  - 2. CA4 31A CA4 21B
  - 3. CA4 31A COB 21A
  - 4. COB 31A COB 21B

- LS 23.0%
  - 5. CA4 31A COB 21B
  - 6. CA4 31A CA4 21B
  - 7. CA4 31A COB 21A
  - 8. COB 31A COB 21B

**Monomodular 17.8%**

- L 6.0%
  - 9. CA4 21B
  - 10. COB 21B
  - 11. CA4 21A

- S 11.0%
  - 12. COB 21B

**Trimodular 14.0%**

- LLL 7.33%
  - 13. CA4 31A CA4 31A COB 21B
  - 14. CA4 31A CA4 31A COB 21B
  - 15. CA4 31A COB 21B COB 21B
  - 16. CA4 31A COB 21B COB 21B

- LLS 0.33%
  - 17. CA4 31A CA4 21B COB 21B
  - 18. CA4 31A COB 21B COB 21B

- LSS 6.33%
  - 19. CA4 31A CA4 21B COB 21B
  - 20. CA4 31A COB 21B COB 21B

**Figure 3.** The frequencies and structures of RCCX modules in the normal Caucasian population. The most common haplotypes are in bold. L, long C4 gene; S, short C4 genes.
monomodular (M/M) and homozygous trimodular (T/T) combinations are less frequent, equaling 2.0 and 3.3%, respectively (Fig. 4A).

Altogether, 17 different diploid combinations of the RCCX structures were detected in the normal population. Among them, LL/LS and LL/LL are most frequent, with 20.7 and 19.3%, respectively. The third most prevalent diploid is LL/L, with a frequency of 12.0%. The other 13 diploid combinations have frequencies varying between 0.67 and 8.67% (Fig. 4C).

It is notable that within the combinations, 90.4% of individuals have at least one chromosome with a bimodular structure. From another view, 31.3% have at least one chromosome with a monomodular structure, and 24.6% have at least one chromosome with a trimodular structure. In total, 69.4% of the diploid genomes have heterozygous combinations of RCCX modules that differ in length by the number of RCCX modules and/or the size of C4 genes.

Number of C4 Genes. The frequency of C4 gene dosage variation is shown in Fig. 4B. Most individuals have four C4 genes (frequency 52.0%) that may be the result of bimodular/bimodular (B/B) or monomodular/trimodular (M/T) diploids. Individuals with three C4 genes in a genome, which are due to the presence of monomodular/bimodular (M/B) combinations, comprise 25.3% of the normal population. Individuals with bimodular/trimodular (B/T) structures have five C4 genes in a genome. They account for 17.3% of the population. An individual with trimodular/trimodular (T/T) structures has six C4 genes, which has a frequency of 3.3%. Only three subjects are found with monomodular/monomodular (M/M) structures, or two C4 genes in a diploid genome, which translates to a frequency of 2.0%.

The Phenotypes and Genotypes of the Complement Components C4A and C4B

Of the 590 C4 genes present in the normal study population, 322 (54.8%) genes encode for the C4A isotype and 275 (45.2%) encode for the C4B isotype. A total of 588 expressed C4 allotypes with considerable polymorphisms were detected. As expected, the allotypes C4A3 and C4B1 are the most common, consisting of 46.6 and 40.1% of the C4 alleles, respectively. The frequencies of the various C4A and C4B allotypes are listed in Table III.

The molecular basis of the C4B Q0 phenotype. A distinct population with the bimodular structures carries two identical C4 isotypic genes on the same chromosome, that is, both C4 genes encode for the C4A (or for the C4B) pro-
proteins. This combination leads to a C4BQ0 (or C4AQ0) phenotype (Table III). Such a phenomenon contributing to C4BQ0 phenotype is confirmed by the presence of the C4A genes using PshAI (or NlaIV) RFLP and the C4-Rg1 using EcoO109I RFLP, and relatively higher intensities of the C4A proteins in allotyping gels and in immunoblots compared with those for C4B and C4-Ch1. This C4A-C4A homoexpression pattern consists of 13.3% of the RCCX haplotypes in the study population. Specifically, 10% of the RCCX haplotypes encode C4A3, C4A3, 2.27% encode C4A3, C4A2. Two individuals (0.67%) appear to contain bimodular RCCX haplotypes coding for C4A3, C4A6 and for C4A4, C4A5, respectively.

Another cause for the C4BQ0 phenotype is due to the presence of monomodular L structure with a single C4A gene. This RCCX structure has a frequency of 5.0%. No C4B mutant gene has been detected in this study population. Therefore, the C4BQ0 phenotype has a combined frequency of 18.3%.

The Molecular Bases of the C4AQ0 Phenotype. 11% of the haplotypes in the study population have monomodular S chromosomes and all of them appear to code for the C4B1 allotype. In addition, 1% of the haplotypes has a monomodular L chromosomes with single C4B genes. Homoeexpression of C4B proteins from bimodular RCCX structures is present at a frequency of 0.67%; the same frequency is observed for the presence of C4A mutant genes with a 2-bp insertion in exon 29. Therefore, the combined frequency of the C4AQ0 phenotype is 13.3% (Table IV).

Together, the partial and complete C4AQ0 and C4BQ0 phenotypes (or deficiencies) have a combined frequency of 31.6% in the normal population. Complete C4A deficiency is only found in one individual, who has homozygous monomodular S (C071; Fig. 2). Complete C4B deficiencies are only found in two individuals, of whom one has homoeexpression of C4A proteins from both chromosomes (C020; Fig. 2) and the other has heterozygous RCCX haplotypes with C4A homoexpression from the bimodular structure and the presence of a single C4A gene from the monomodular L structure (C033; data not shown).

C4 and CYP21 G enes in RCCX T rimodular H aplotypes. As listed in Table IV, 42 trimodular haplotypes were detected in the normal subjects. Overall, 7.33% of the RCCX haplotypes contain two C4A and one C4B gene, and 6.67% contain one C4A and two C4B genes. The LLL structures more frequently code for C4A-C4A-C4B, whereas the LSS structures more frequently code for C4A-C4B-C4B.

In the same population, the CYP21A-CYP21A-CYP21B configuration was found in 37 haplotypes (12.3%), and the CYP21A-CYP21B-CYP21B configuration was found in 5 haplotypes (1.67%). In other words, 88.1% of the trimodular chromosomes have two CYP21A genes and one CYP21B gene.

**Table III.** Molecular Bases for Partial Deficiencies of C4A and C4B in the Normal Caucasian Population

<table>
<thead>
<tr>
<th>C4AQ0</th>
<th>C4BQ0</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>Percent*</td>
<td>n</td>
</tr>
<tr>
<td>Gene deletion</td>
<td>36</td>
<td>12</td>
</tr>
<tr>
<td>Mono-L</td>
<td>3</td>
<td>1.0</td>
</tr>
<tr>
<td>Mono-S</td>
<td>33</td>
<td>11.0</td>
</tr>
<tr>
<td>Gene mutation</td>
<td>2</td>
<td>0.67</td>
</tr>
<tr>
<td>Gene conversion</td>
<td>2</td>
<td>0.67</td>
</tr>
<tr>
<td>C4A-C4A</td>
<td>(40)</td>
<td>(13.3)</td>
</tr>
<tr>
<td>A3-A2</td>
<td>(8)</td>
<td>(2.67)</td>
</tr>
<tr>
<td>C4B-C4B</td>
<td>(2)</td>
<td>(0.67)</td>
</tr>
<tr>
<td>Total</td>
<td>40</td>
<td>13.3</td>
</tr>
</tbody>
</table>

*Frequency in haplotypes.

**Table IV.** C4A and C4B Genes in Trimodular RCCX Structures

<table>
<thead>
<tr>
<th>C4A-C4A-C4B</th>
<th>C4A-C4B-C4B</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>Percent*</td>
</tr>
<tr>
<td>LLL</td>
<td>14</td>
</tr>
<tr>
<td>LSS</td>
<td>7</td>
</tr>
<tr>
<td>LLS</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>22</td>
</tr>
</tbody>
</table>

*Frequency in haplotypes.
Table V. RCCX Structures, CYP21A and CYP21B Genes in Patients with Classical CAH

<table>
<thead>
<tr>
<th>RCCX</th>
<th>Patients (n=44)</th>
<th>Normal subjects (n=300)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Percent*</td>
<td>n</td>
</tr>
<tr>
<td>Mono-L</td>
<td>11</td>
<td>25.0</td>
<td>18</td>
</tr>
<tr>
<td>Mono-S</td>
<td>0</td>
<td>0</td>
<td>33</td>
</tr>
<tr>
<td>LL</td>
<td>16</td>
<td>36.4</td>
<td>138</td>
</tr>
<tr>
<td>LS</td>
<td>13</td>
<td>29.5</td>
<td>70</td>
</tr>
<tr>
<td>LLL</td>
<td>2</td>
<td>4.5</td>
<td>21</td>
</tr>
<tr>
<td>LSS</td>
<td>2</td>
<td>4.5</td>
<td>19</td>
</tr>
<tr>
<td>CYP21A in mono-L</td>
<td>9</td>
<td>20.5</td>
<td>2</td>
</tr>
<tr>
<td>CYP21A-CYP21A</td>
<td>11</td>
<td>22.7</td>
<td>4</td>
</tr>
<tr>
<td>CYP21A-CYP21A-CYP21B</td>
<td>4</td>
<td>9.1</td>
<td>37</td>
</tr>
<tr>
<td>CYP21A-CYP21B-CYP21B</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>

Number of CAH patients studied: 22 (number of chromosome 6: 44); total number of C4 genes detectable: 81; frequency of long C4 genes: 79.0% (normal 76.2%); frequency of short C4 genes: 21% (normal 23.7%).

*N Frequency in haplotypes.

Discussion

The high frequency and complexity of complement component C4 gene size, number, and modular variations with flanking genes RP, CYP21, and TNX in the Caucasian population are described. The RCCX modules were investigated through genetic analysis using six different probes for diagnostic RFLP analyses to define genomic structures including the number and size of genes, the presence of C4A versus C4B, and the association of the C4 genes with Rg1 and Ch1 antigenic determinants. Combined with protein typing data, accurate assignments of genotype-phenotype relationships were made. The etiology of C4 null alleles through gene deletion, gene duplication, and mutant genes was defined.

The number of C4 gene loci present in humans was a controversial topic until O’Neill et al. (37), Awdeh and Alper (35), and Rittner and colleagues (38) proposed a two-loci theory with C4A and C4B genes. Thereafter, individuals with one or three C4 genes have been documented, which were largely presumed to be caused by secondary gene deletions and gene duplications (14, 16, 39). Based on C4 protein typing data, it was estimated that in the Caucasian population, the frequency for the half-null allele for C4A was 9.5–14.5%, and the frequency for the half-null allele of C4B was 16–19%. “Homoduplication” of C4A or C4B, defined as the presence of three C4 genes or two C4 genes coding for the same C4A or C4B isotypes, was 1–2% (23). The current study firmly establishes a “1-2-3 loci” concept for C4 and CYP21 in the Caucasian population.

We have observed an apparent, partial C4A deficiency at a frequency of 13.3% and partial C4B deficiency at a frequency of 18.3%. This extremely high degree of apparent, partial protein deficiencies in the population is predominantly the result of the presence of monomodular structures and homoexpression of identical C4A or C4B isotypes from bimodular structures. The C4A mutant gene due to the 2-bp insertion in codon 1213 (in exon 29) was detected in only two individuals of the normal study population.

The frequency of three C4 gene haplotypes (or trimodular RCCX structures) was underestimated in previous studies. As stated above, we identified that 14.0% of the RCCX structures in chromosome 6 are trimodular. The trimodular structures are predominantly present as LLL and LSS; the former more frequently codes for C4A3-A3-B1, and the latter, C4A3-B1-B1. These trimodular configurations from bimodular RCCX structures (Fig. 5, lanes 4 and 6) account for 22.7% of the disease haplotypes. Overall, 43.2% of the RCCX haplotypes in the CAH population do not contain a CYP21B gene. Compared with the normal group, the CAH population has a striking increase of monomodular RCCX with a long C4 gene (25 vs. 6%; P < 0.001) and a total absence of the monomodular RCCX with a short C4 gene (mono-S).

Figure 5. RCCX modular variations in classical CAH patients. (A) TaqI RFLP on seven selected CAH patients to show the RCCX diploid combinations. The 8-digit number above each lane represents the relative dosage of the 7.0-, 6.4-, 6.0-, and 5.4-kb fragments for RP-C4, and the 3.7-, 3.2-, 2.5-, and 2.4-kb fragments for CYP21B, CYP21A, TNXB, and TNXA, respectively, as shown in Table I. (B) PsIAI RFLP to determine the C4A and C4B gene dosage in the corresponding patients. (C) Interpretation of the RCCX structures and the relative gene dosage of CYP21A and CYP21B, and of C4A and C4B.

Number of CAH patients studied: 22 (number of chromosome 6: 44); total number of C4 genes detectable: 81; frequency of long C4 genes: 79.0% (normal 76.2%); frequency of short C4 genes: 21% (normal 23.7%).

* Frequency in haplotypes.
tions would manifest the false, partial C4A or C4B deficiency phenotypes in C4 allotyping gels if not coupled with a systematic analysis of the RCCX genomic structures. In other words, the protein allotyping data alone are not enough to define the etiologic mechanisms behind relationships between complement C4A or C4B deficiency and autoimmune disease. To understand the relationship of C4A and C4B polymorphism in autoimmunity, it is important to clearly identify if the process is related to the overexpression of C4, which may exacerbate the disease process through excessive complement activation, or the underexpression of C4, resulting in impaired immunoclearance. There is a high likelihood of overexpression in one of the C4 isotypes due to the high frequency of trimodular structures and/or homoexpression of C4A or C4B.

In this study, it is found that the relative titers of the C4A and C4B proteins in a blood plasma sample closely correspond to the relative C4A and C4B gene dosage in the same individual. The variation of C4A and C4B protein titers in the population had been investigated by many laboratories, using Rg1 and Ch1 mAbs in enzyme-linked immunosorbent assays (40–42). The overall complement C4 levels in an individual may be related to the physiological state and the status of the RCCX modules. A revisit of the quantitation of the plasma or serum C4A and C4B levels with respect to the C4 gene dosage will be necessary for an accurate account of the expression of each C4 gene and its protein levels in the circulation.

Although in 90.4% of the Caucasian population the diploid genome carries two C4 genes in one of the chromosome 6, 55.9% of the same population carries one or three C4 genes in the other chromosome 6. These trimodular and monomodular structures increase the continuing likelihood of atypical genetic recombinations or unequal crossovers. The misalignments that occur result in diseases such as CAH where, as with the complement C4 genes, there is a high sequence homology (98.7%) between CYP21B and its pseudogene CYP21A. This pseudogene carries 58 point mutations, 8 mini-insertions and two mini-deletions in the 5.14-kb sequence that include 5’ and 3’ flanking regions (43–46). The heterozygosity of a bimodular RCCX with a monomodular or a trimodular structure

![Diagram](image-url)

**Figure 6.** Alignments of the heterozygous RCCX length variants during meiosis to generate CAH disease haplotypes (A and B), or to facilitate appropriate pairings by the mono-S haplotype (C). For A, a pairing of a trimodular LLL structure with a mono-L structure may give a similar result. For C-3, a pairing at the telomeric end may not allow a correct alignment between the CYP21 genes from the homologous chromosomes and therefore may reduce the exchange of sequences.
is likely to create the interchromosomal pairing of a CYP21B gene with a CYP21A gene during meiosis (Fig. 6). Exchange of genetic information between the CYP21B and CYP21A by intergenic recombination may reduce or abrogate the functionality of the CYP21B gene, depending on the site and extent of exchanges.

Specifically, the CYP21B gene in the monomodular-long haplotype (mono-L) would be more vulnerable to crossover or conversion in heterozygous RCCX individuals due to its intrinsic misalignments with the CYP21A in a bimodular or trimodular chromosome (Fig. 6 A). In this study, it is found that 20.5% of the salt-losing CAH patients have mono-L with CYP21A, a phenomenon similar to that described previously (46). It is also found that 22.7% of the CAH haplotypes have the CYP21A-CYP21A configuration in bimodular structures. This disease haplotype could be the result of recombinations between a trimodular chromosome with CYP21A-CYP21A-CYP21B configuration and a regular bimodular chromosome (at or after the second CYP21A gene of the trimodular chromosome; Fig. 6 B). Intriguingly, the mono-S haplotype, which is present in the normal subjects at a frequency of 11% and common in the autoimmune disease patients, is totally absent in the CAH population, suggesting a protective role of this haplotype against the disease. The mono-S haplotype probably enhances the correct CYP21B/CYP21B pairings at the centromeric end of the MHC among heterozygous RCCX length variants (Fig. 6, C-1 and C-2) and discourages alignments of the RCCX at the telomeric end due to a greater degree of dissimilarity (Fig. 6 C-3).

The physiological advantage of this new paradigm of genetic diversity created by length polymorphisms is still a matter of investigation. Although it has not been critically examined, it is probable that these RCCX length polymorphisms may influence the linkage disequilibrium phenomenon of the HLA class I, II, and III genes (47, 48). In terms of immune defense, this may be a way to create a variety of individuals with varying quantities of complement C4A or C4B proteins, and to homogenize polymorphic sequences among different C4 genes. It is observed here that there are slightly more C4 loci coding for C4A isoforms (C4A, 54%; C4B, 46%), even though 11% of the chromosome 6 are monomodular with a single C4B gene. This may infer a selection pressure to keep (if not to increase) the frequency of C4A genes in the population. This could be for greater efficiency in handling immune complexes, binding to complement receptor CR1, and inhibition of immunoprecipitation, as C4A plays a larger functional role with this regard than C4B (5, 49).

We are grateful to the blood donors who made this study possible. We are indebted to the Dr. John Moulds for the gift of the C4 mAbs and instruction of C4 allotyping techniques.

This work was supported by grants from the National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health (R01 AR 43969), the Children’s Research Institute, Columbus, OH (21999), the March of Dimes Birth Defects Foundation (FY 95-1087, Basil O’Conner Starter Scholar Research Award), the Central Ohio Diabetes Association, and the Pittsburgh Supercomputing Center through the National Institutes of Health Center for Research Resources Cooperative Agreement (1P41 RR-06009).

Submitted: 2 November 1999
Revised: 17 March 2000
Accepted: 31 March 2000

References


