

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 concurs 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).

Address correspondence to C. Yung Yu, Rm. W208, Children's Research Institute, Department of Pediatrics, The Ohio State University, 700 Children's Dr., Columbus, OH 43205-2696. Phone: 614-722-2821; Fax: 614-722-2774; E-mail: cyu@chi.osu.edu

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)¹ 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 PCPVLD and C4B has LSPVIH (11). The Rg1 and Ch1 are the results of two amino acid substitutions at residues 1188–1191. The Rg1 epitope has VDLL, 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 *RP*, steroid 21-hydroxylase *CYP21*, and extracellular matrix protein tenascin *TNX*. These four genes form a genetic unit designated as the RCCX module (17, 18). The duplication of RCCX in a bimodular structure generates an additional gene for complement component C4, but the other three concurrently duplicated constituents located between the two C4 genes are nonfunctional. Each human C4 gene contains 41 exons, and the gene size shows a dichotomous size variation between 21 and 14.6 kb. The long gene is due to the integration of the endogenous retrovirus HERV-K(C4) into intron 9 (19–22). The frequency of this HERV-K(C4) in the C4 genes has not been defined although it appears that this endogenous retrovirus is always present in the first C4 locus that usually codes for a C4A protein. The physiological significance of this family of endogenous retroviruses has not been elucidated.

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 comple-

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

¹Abbreviations 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.

DNA Probes. The positions of the genomic DNA in the MHC complement gene cluster (MCGC) to which the probes hybridized are illustrated in Fig. 1. The DNA fragments for probes A and C were derived from cloned DNA fragments in plasmids; the others were generated by PCR using the appropriate primers and DNA templates. The probes were as follows: probe A, a 1.1-kb cDNA fragment of *RP1*, or a 600-bp *NheI*-*EcoRI* fragment corresponding to the 3' end of the DNA sequence present in both *RP1* and *RP2*; probe B, a 0.8-kb fragment corresponding to exons 6–9 of the *C4* gene, amplified by PCR using *E65* and *E93* as primers and *cos 3A3* as template; probe C, also referred to as *P_B*, a 926-bp *BamHI* fragment of the *C4d* region from λ JM-2a; probe D, a 1.1-kb fragment corresponding to exons 28–31 of *C4*, amplified by primers *E285* and *E313* and *cos 3A3* as template; probe E, a 757-bp fragment corresponding to exons 4–7 of *CYP21*, amplified from cosmid DNA *cos 4A3*, using primers *21A5* and *21A3*; and probe F, a 500-bp fragment corresponding to exons 35–37 of *TNXA*, amplified using primers *RDX-5* and *RDX-3* and *cos 4A3* as the DNA template (17, 33).

Complement C4 Allotyping and Immunoblot Analysis. Peripheral blood plasma from EDTA samples was used to test for allotypic polymorphisms of complement C4 following standard procedures (34–36). Some of the C4A and C4B phenotypes were confirmed with Western immunoblot analysis using anti-Ch1 and anti-Rg1 mAbs (anti-C4B; a gift from Dr. Joann Moulds, University of Texas, Houston, TX) at a dilution of 1:5,000 and 1:1,000, respectively. Immune complexes were detected by the chemiluminescence method using ECL Plus reagents (Amersham Pharmacia Biotech).

Detection of C4 Mutants. Sequence-specific PCRs were used to detect the known C4 mutations in samples whose C4 protein products could not be appropriately accounted for. The PCR primers for detection of the 2-bp insertion in exon 29 were: (A-down) AGG ACC CCT GTC CAG TGT TAG AC, (B-down) AGGACCTCTCTCCAGTGATACAT, and (E29-INS) GCT CTG AGA ACC AGT GAC TAG AG. PCR conditions were 1 cycle at 94°C for 5 min; 25 cycles at 94°C for 30 s, 65°C for 45 s, and 72°C for 1 min; and 1 cycle at 72°C for 10 min.

Results

An intensive study of the molecular organization of 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 the ma-

ajor 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 *TaqI* 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 *RP1-C4L-CYP21A-TNXA-RP2-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 *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*.

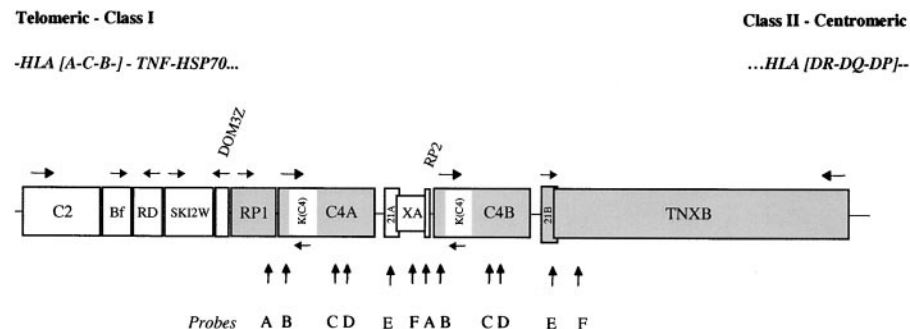


Figure 1. A schematic diagram of the human MHC complement gene cluster. The functional genes of the RCCX modules are highlighted. A horizontal arrow represents the transcriptional orientation of a gene; vertical arrows represent locations of the DNA hybridized to probes A–F. K(C4), endogenous retrovirus HERV-K(C4); 21A, *CYP21A*; 21B, *CYP21B*; XA, *TNXA*.

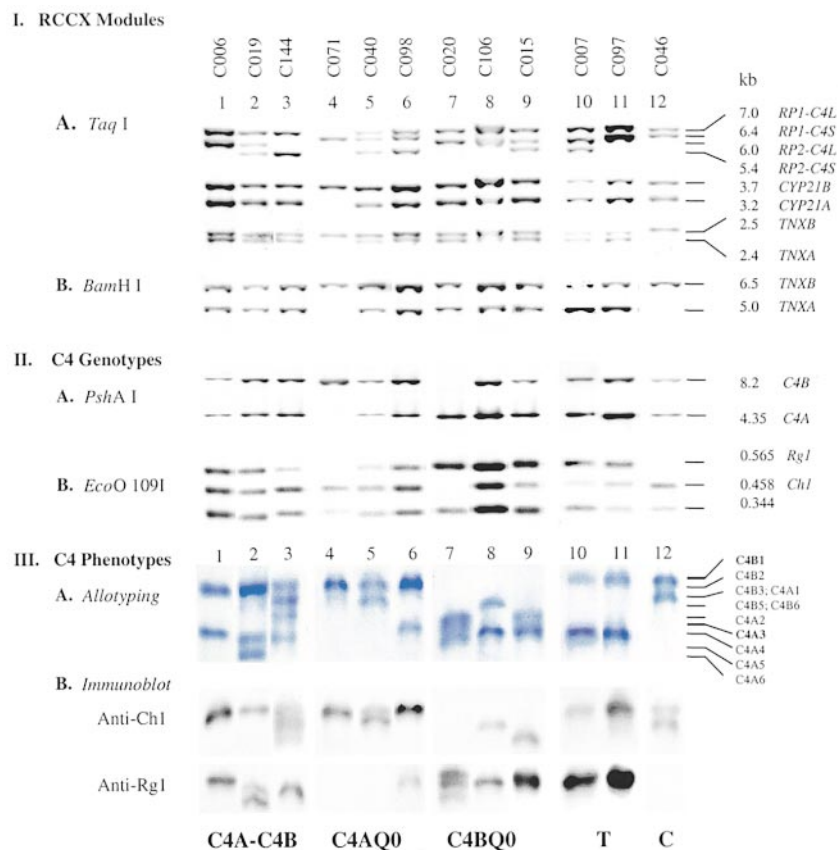


Figure 2. Genotypic and phenotypic analyses of the RCCX modules in selected individuals. (I) TaqI and BamHI RFLPs, to define the RCCX modular structure of each chromosome. Probes A, E, and F were used for TaqI RFLP; probe E was used for BamHI RFLP. (II) Detection of the *C4A* and *C4B* genes and *C4* genes associated with Rg1 or Ch1 by PshAI (A) and EcoO1091 (B) RFLPs with probes C and D, respectively. (III) Detection of *C4A* and *C4B* allotypes by immunofixation (A) and immunoblot analysis (B), respectively. T, trimodular; C, CAH carrier.

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 EcoO1091 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 *C4A12 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 EcoO1091 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

Table I. Relative Gene Dosage of the RCCX Modules from Selected Individuals

	TaqI				BamHI		PshAI		EcoO109I*		RCCX modules	Probable C4 haplotypes					
	RP-C4 [‡]				XB	XA	C4B	C4A	Rg1	Ch1							
	7.0	6.4	6.0	5.4	3.7	3.2	2.5	2.4	6.5	4.9	8.2	4.35	565	457			
I																	
1. C006	2 [§]	0	2	0	2	2	2	2	2	2	2	2	2	2	2	LL/LL	A3 B1/A3 B1
2. C019	2	0	1	1	2	2	2	2	2	2	2	2	2	2	2	LL/LS	A4 B1/A6 B1
3. C144	2	0	0	2	2	2	2	2	2	2	2	2	1	3	3	LS/LS	A4 B2/A12 (Ch1) B1
II																	
4. C071	0	2	0	0	2	0	2	0	2	0	2	0	0	2	2	S/S	AQ0 B1/AQ0 B1
5. C040	1	1	0	1	2	1	2	1	2	1	2	1	1	2	2	LS/S	AQ0 B2/AQ0 B1
6. C098	1	1	0	1	2	1	2	1	2	1	2	1	1	2	2	LS/S	A3 B1/AQ0 B1
III																	
7. C020	2	0	2	0	2	2	2	2	2	2	0	4	4	0	0	LL/LL	A3 A2/A2 A5
8. C106	2	0	1	0	2	1	2	1	2	1	1	2	2	1	1	LL/L	A3 B3/A3 BQ0
9. C015	2	0	1	1	2	2	2	2	2	2	1	3	3	1	1	LL/LS	A3 A3/A3 B6
IV																	
10. C007	2	0	2	1	2	3	2	3	2	3	2	3	3	2	2	LLL/LS	A3 A3 B1/A3 B1
11. C097	2	0	4	0	2	4	2	4	2	4	2	4	4	2	2	LLL/LLL	A3 A3 B1/A3 A3 B1
V																	
12. C046	1	1	0	0	1	1	2	0	2	0	1	1	0	2	2	L/S	A1 (Ch1) BQ0/AQ0 B1

21A, CYP21A; 21B, CYP21B; XA, TNXA; XB, TNXB.

*Restriction fragment size in bp.

[‡]RP1-C4(L), 7 kb; RP1-C4(S), 6.4 kb; RP2-C4(L), 6 kb; RP2-C4(S), 5.4 kb (sizes of all fragments in kb except for EcoO109I).

[§]Relative gene dosage.

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. The 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-

iciency 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 C097 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 *TNXA* is greater than 2.5-kb *TNXB*. In addition, the 5.0-kb BamHI fragment for *TNXA* 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 *TNXA* suggest that one of the chromosome 6 has a trimodular RCCX structure with three long C4 genes. The presumed RCCX haplotypes for C007 are trimodular *RP1-C4A3(L)-CYP21A-TNXA-RP2-C4A3(L)-CYP21A-TNXA-RP2-C4B1(L)-CYP21A-TNXB* and bimodular *RP1-C4A3(L)-CYP21A-TNXA-RP2-C4B1(S)-CYP21B-TNXB*.

Results of the RFLP studies for C097 reveal homozygous 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 C097 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-C4B-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 allotyping gel and immunoblot experiments confirmed the presence of two C4 proteins expressing the Ch1 epitope. The predicted C4 haplotypes are A_{Q0} B1/A1 B_{Q0}. Similar to the C4A12 in C0144, C4A1 has the reverse association with Ch1. The C4A_{Q0} or C4B_{Q0} in each case is due to the deletion of an RCCX module.

RCCX Modular Variations in the Normal Caucasian Population

The study population with 150 normal Caucasian females was analyzed as described above. From the extensive RFLP experiments, C4 allotyping and immunoblot analyses, the frequencies of the long and short C4 genes, the monomodular, bimodular, and trimodular RCCX structures in haplotypes and in diploids, and the bases of the C4A and C4B deficiencies are determined.

Long and Short C4 Genes. The presence of the long and short C4 genes is determined by two independent RFLP experiments. From the TaqI Southern blots, the long C4 genes are designated with the 7.0- and 6.0-kb fragments, the short C4 genes with the 6.4- and 5.4-kb fragments. From the BamHI Southern blots hybridized with probe B (Fig. 1), the long C4 genes are indicated by the presence of the 4.8-kb fragments and the short C4 genes by the 3.3-kb fragments (data not shown). Among the 590 complement C4 genes present in the study population, there are 450 (76.2%) long genes and 140 (23.7%) short genes. This result implies that 3/4 of the population has the endogenous retrovirus HERV-K(C4) in their C4 genes and 1/4 does not (Table II).

RCCX Modules. The presence of a bimodular RCCX haplotype is determined by the concurrent presence and the equimolar ratio of the *RP1-C4(L)* (7.0 kb) with the *RP2-C4(L)* (6.0 kb) (i.e., LL), or the *RP2-C4(S)* (5.4 kb) fragments (i.e., LS) in the TaqI RFLP. It is also indicated

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

C4 genotypes					
No. of individuals studied	150				
(No. of chromosome 6)	(300)				
Total no. of C4 genes present	590				
Frequency of long C4 genes	76.2%				
Frequency of short C4 genes	23.7%				
Phenotypes of C4A and C4B					
Allotype	n	Percent*	Allotype	n	Percent*
A1	1	0.17	B1	236	40.1
A12	1	0.17	B13	1	0.17
A2	17	2.89	B2	25	4.25
A3	274	46.6	B3	2	0.34
A4	11	1.87	B6	2	0.34
A5	3	0.51			
A6	15	2.55			
Total	322	54.8	Total	266	45.2

*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%.

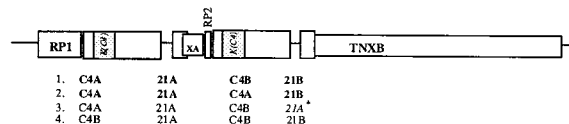
In contrast to the monomodular RCCX haplotypes, the trimodular structure is defined by relatively higher intensities of the restriction fragments for *RP2-C4(L)* and/or *RP-C4(S)* compared with that for the *RP1-C4(L)* in a TaqI RFLP. Similarly, the ratio of *RP1* to *RP2* and of *TNXB* to *TNXA* is <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.

Diploid RCCX Combinations. The various combinations of RCCX haplotypes in diploids are shown in Fig. 4 (A and C). In brief, the most common combination is bimodular/bimodular (B/B), as it is present in 48% of the study population. The frequencies for monomodular/bimodular (M/B) and the bimodular/trimodular (B/T) combinations are 25.3 and 17.3%, respectively. Homozygous

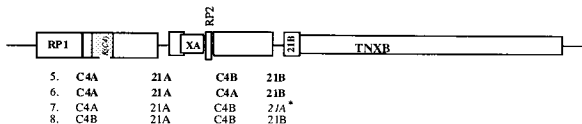
RCCX MODULES

Bimodular 69.0%

LL 46.0%

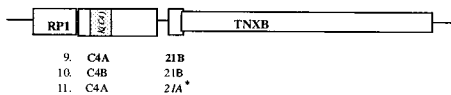


LS 23.0%

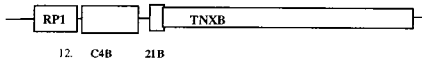


Monomodular 17.0%

L 6.0%

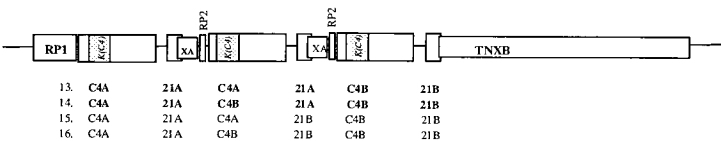


S 11.0%

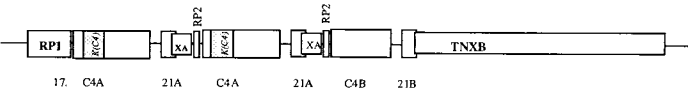


Trimodular 14.0%

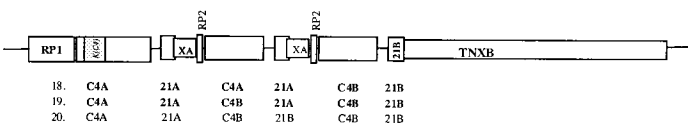
LLL 7.33%



LLS 0.33%



LSS 6.33%



10 20 kb

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.

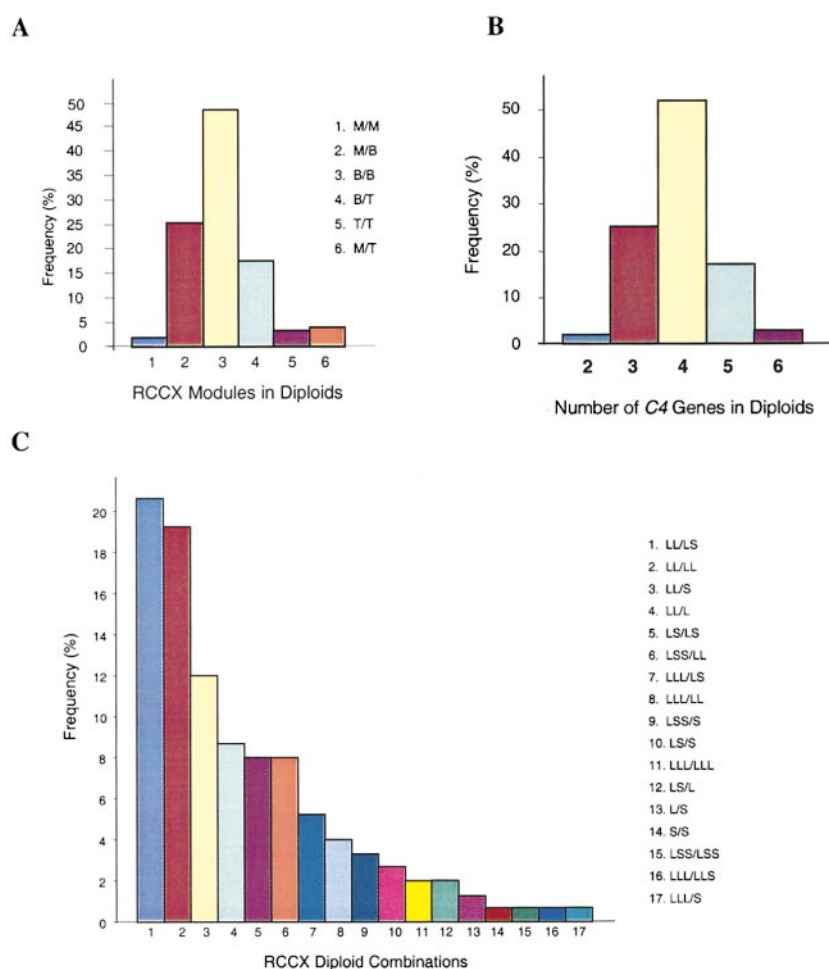


Figure 4. Diploid combinations of the RCCX modules in the Caucasian population. (A) Frequencies of RCCX modular combinations based on the number of modules. (B) Variations in the number of *C4* genes. (C) Frequencies of the RCCX modules based on the number of modules and size of *C4* genes.

monomodular (M/M) and homozygous trimodular (T/T) combinations are less frequent, equaling 2.0 and 3.3%, respectively (Fig. 4 A).

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/S, with a frequency of 12.0%. The other 13 diploid combinations have frequencies varying between 0.67 and 8.67% (Fig. 4 C).

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 the *C4* genes.

Number of *C4* Genes. The frequency of *C4* gene dosage variation is shown in Fig. 4 B. 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* allotypes, respectively. The frequencies of the various C4A and C4B allotypes are listed in Table III.

The Molecular Bases of the C4BQ0 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-

teins. 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. Homoexpression 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 homoexpression 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 Genes in RCCX Trimodular Haplotypes. 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.

RCCX Structures and CYP21 Genes in Patients with Classical CAH

Among the 150 normal subjects, two individuals have a heterozygous deletion of the *CYP21B* gene in RCCX monomodular structures (frequency in haplotypes 0.67%). In addition, four individuals have two *CYP21A* genes from the bimodular RCCX haplotypes (frequency 1.33%). These six individuals are carriers of the classical form of CAH.

As an example to illustrate the impact of RCCX modular variations in MHC-associated diseases, we studied the 22 patients with the classical salt-losing form of CAH. The results are listed in Table V. Seven typical diploid combinations of the RCCX structures from the CAH patients are shown in Fig. 5.

Monomodular, bimodular, and trimodular RCCX structures are detectable in the CAH population. However, monomodular structures with a long *C4* gene (mono-L) with the presence of *CYP21A* (i.e., deletion of *CYP21B*; Fig. 5, lanes 1 and 2) have a haplotype frequency of 20.5% ($P < 0.001$). Gene conversions with two *CYP21A* pseudo-

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

	C4AQ0		C4BQ0		Total	
	<i>n</i>	Percent*	<i>n</i>	Percent*	<i>n</i>	Percent*
Gene deletion	36	12	15	5.	51	17
<i>Mono-L</i>	3	1.0	15	5.0	18	6
<i>Mono-S</i>	33	11.0	0	0	33	11
Gene mutation	2	0.67	0	0	2	0.67
Gene conversion	2	0.67	40	13.3	42	14.0
<i>C4A-C4A</i>			(40)	(13.3)		
<i>A3-A3</i>			(30)	(10.0)		
<i>A3-A2</i>			(8)	(2.67)		
Others			(2)	(0.67)		
<i>C4B-C4B</i>	(2)	(0.67)				
Total	40	13.3	55	18.3	95	31.6

*Frequency in haplotypes.

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

	<i>C4A-C4A-C4B</i>		<i>C4A-C4B-C4B</i>	
	<i>n</i>	Percent*	<i>n</i>	Percent*
LLL	14	4.67	8	2.67
LSS	7	2.33	12	4.0
LLS	1	0.33	0	0
Total	22	7.33	20	6.67

*Frequency in haplotypes.

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

RCCX	Patients (n = 44)		Normal subjects (n = 300)		P
	n	Percent*	n	Percent*	
Mono-L	11	25.0	18	6.0	<0.001
Mono-S	0	0	33	11.0	<0.001
LL	16	36.4	138	46.0	
LS	13	29.5	70	23.3	0.23
LLL	2	4.5	21	7.0	0.368
LSS	2	4.5	19	6.33	
CYP21A in mono-L	9	20.5	2	0.67	<0.001
CYP21A-CYP21A	11	22.7	4	1.33	<0.001
CYP21A-CYP21A-CYP21B	4	9.1	37	12.3	
CYP21A-CYP21B-CYP21B	0	0	5	1.67	

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.

genes 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).

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 allotyping 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 Cauca-

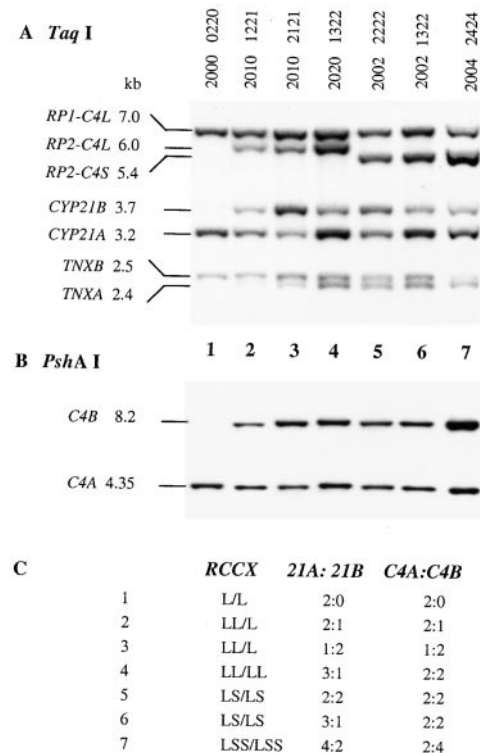


Figure 5. RCCX modular variations in classical CAH patients. (A) TaqI RFLP on seven selected CAH patients to show the RCCX diploid combinations. The 8-digit number above each lane represents the relative dosage of the 7.0-, 6.4-, 6.0-, and 5.4-kb fragments for RP-C4, and the 3.7-, 3.2-, 2.5-, and 2.4-kb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

sian 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 configura-

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 disease states 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

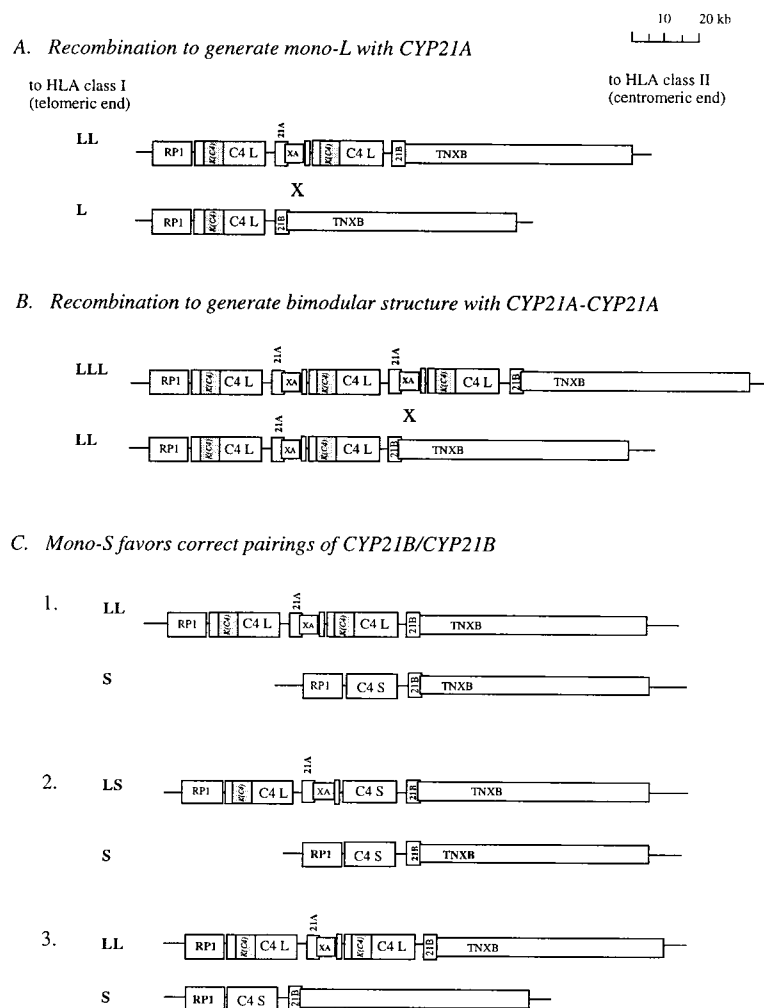


Figure 6. Alignments of the heterozygous RCCX length variants during meiosis to generate CAH disease haplotypes (A and B), or to facilitate appropriate pairings by the mono-S haplotype (C). For A, a pairing of a (trimodular) LLL structure with a mono-L structure may give a similar result. For C-3, a 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 isotypes (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. Joann 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 AR43969), the Children's Research Institute, Columbus, OH (21999), the March of Dimes Birth Defects Foundation (FY95-1087, 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 RR06009).

Submitted: 2 November 1999

Revised: 17 March 2000

Accepted: 31 March 2000

References

1. Klein, J., N. Takahata, and F.J. Ayala. 1993. MHC polymorphism and human origins. *Sci. Am.* 269:78-83.
2. Dawkins, R., C. Leelayuwat, S. Gaudieri, G. Tay, J. Hui, S. Cattley, P. Martinez, and J. Kulski. 1999. Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses and disease. *Immunol. Rev.* 167:275-304.
3. Yu, C.Y. 1998. Molecular genetics of the human MHC complement gene cluster. *Exp. Clin. Immunogenet.* 15:213-230.
4. Mauff, G., B. Luther, P.M. Schneider, C. Rittner, B. Strandmann-Bellinghausen, R. Dawkins, and J.M. Moulds. 1998. Reference typing report for complement component C4. *Exp. Clin. Immunogenet.* 15:249-260.
5. Law, S.K.A., A.W. Dodds, and R.R. Porter. 1984. A comparison of the properties of two classes, C4A and C4B, of the human complement component C4. *EMBO (Eur. Mol. Biol. Organ.) J.* 3:1819-1823.
6. Dodds, A.W., X.-D. Ren, A.C. Willis, and S.K.A. Law. 1996. The reaction mechanism of the internal thioester in the human complement component C4. *Nature.* 379:177-179.
7. Isenman, D.E., and J.R. Young. 1984. The molecular basis for the differences in immune hemolysis activity of the Chido and Rodgers isotypes of human complement component C4. *J. Immunol.* 132:3019-3027.
8. Carroll, M.C., D.M. Fathallah, L. Bergamaschini, E.M. Alicot, and D.E. Isenman. 1990. Substitution of a single amino acid (aspartic acid for histidine) converts the functional activity of human complement C4B to C4A. *Proc. Natl. Acad. Sci. USA.* 87:6868-6872.
9. Giles, C.M., B. Uring-Lambert, J. Goetz, G. Hauptmann, A.H.L. Fielder, W. Ollier, C. Rittner, and T. Robson. 1988. Antigenic determinants expressed by human C4 allotypes: a study of 325 families provides evidence for the structural antigenic model. *Immunogenetics.* 27:442-448.
10. Yu, C.Y., and R.D. Campbell. 1987. Definitive RFLPs to distinguish between the human complement C4A/C4B isotypes and the major Rodgers/Chido determinants: application to the study of C4 null alleles. *Immunogenetics.* 25:383-390.
11. Yu, C.Y., K.T. Belt, C.M. Giles, R.D. Campbell, and R.R. Porter. 1986. Structural basis of the polymorphism of human complement component C4A and C4B: gene size, reactivity and antigenicity. *EMBO (Eur. Mol. Biol. Organ.) J.* 5:2873-2881.
12. Yu, C.Y., R.D. Campbell, and R.R. Porter. 1988. A structural model for the location of the Rodgers and the Chido antigenic determinants and their correlation with the human complement C4A/C4B isotypes. *Immunogenetics.* 27:399-405.
13. Dodds, A.W., S.-K.A. Law, and R.R. Porter. 1986. The purification and properties of some less common allotypes of the fourth component of human complement. *Immunogenet-*

- ics. 24:279–285.
14. Carroll, M.C., A. Palsdottir, K.T. Belt, and R.R. Porter. 1985. Deletion of complement C4 and steroid 21-hydroxylase genes in the HLA class III region. *EMBO (Eur. Mol. Biol. Organ.) J.* 4:2547–2552.
 15. Carroll, M.C., R.D. Campbell, D.R. Bentley, and R.R. Porter. 1984. A molecular map of the human major histocompatibility complex class III region linking complement genes C4, C2 and factor B. *Nature.* 307:237–241.
 16. Schneider, P.M., M.C. Carroll, C.A. Alper, C. Rittner, A.S. Whitehead, E.J. Yunis, and H.R. Colten. 1986. Polymorphism of human complement C4 and steroid 21-hydroxylase genes. Restriction fragment length polymorphisms revealing structural deletions, homoduplications, and size variants. *J. Clin. Invest.* 78:650–657.
 17. Yang, Z., A.R. Mendoza, T.R. Welch, W.B. Zipf, and C.Y. Yu. 1999. Modular variations of HLA class III genes for serine/threonine kinase RP, complement C4, steroid 21-hydroxylase CYP21 and tenascin TNX (RCCX): a mechanism for gene deletions and disease associations. *J. Biol. Chem.* 274:12147–12156.
 18. Shen, L.M., L.C. Wu, S. Sanlioglu, R. Chen, A.R. Mendoza, A. Dangel, M.C. Carroll, W. Zipf, and C.Y. Yu. 1994. Structure and genetics of the partially duplicated gene RP located immediately upstream of the complement C4A and C4B genes in the HLA class III region: molecular cloning, exon-intron structure, composite retroposon and breakpoint of gene duplication. *J. Biol. Chem.* 269:8466–8476.
 19. Yu, C.Y. 1991. The complete exon-intron structure of a human complement component C4A gene: DNA sequences, polymorphism, and linkage to the 21-hydroxylase gene. *J. Immunol.* 146:1057–1066.
 20. Dangel, A.W., A.R. Mendoza, B.J. Baker, C.M. Daniel, M.C. Carroll, L.-C. Wu, and C.Y. Yu. 1994. The dichotomous size variation of human complement C4 gene is mediated by a novel family of endogenous retroviruses which also establishes species-specific genomic patterns among Old World primates. *Immunogenetics.* 40:425–436.
 21. Chu, X., C. Rittner, and P.M. Schneider. 1995. Length polymorphism of the human complement component C4 gene is due to an ancient retroviral integration. *Exp. Clin. Immunogenet.* 12:74–81.
 22. Tassabehji, M., T. Strachan, M. Anderson, R.D. Campbell, S. Collier, and M. Lako. 1994. Identification of a novel family of human endogenous retroviruses and characterization of one family member, HERV-K(C4), located in the complement C4 gene cluster. *Nucleic Acids Res.* 22:5211–5217.
 23. Hauptmann, G., J. Goetz, B. Uring-Lambert, and E. Grosshans. 1986. Component deficiencies. *Prog. Allergy.* 39:232–249.
 24. Braun, L., P.M. Schneider, C.M. Giles, J. Bertrams, and C. Rittner. 1990. Null alleles of human complement C4. Evidence for pseudogenes at the C4A locus and for gene conversion at the C4B locus. *J. Exp. Med.* 171:129–140.
 25. Palsdottir, A., A. Arnason, R. Fosdhal, and O. Jensson. 1987. Gene organization of haplotypes expressing two different C4A allotypes. *Hum. Genet.* 76:220–224.
 26. Barba, G., C. Rittner, and P.M. Schneider. 1993. Genetic basis of human complement C4A deficiency: detection of a point mutation leading to nonexpression. *J. Clin. Invest.* 91:1681–1686.
 27. Lokki, M.-L., A. Circolo, P. Ahokas, K.L. Rupert, C.Y. Yu, and H.R. Colten. 1998. Deficiency of complement protein C4 due to identical frameshift mutations in the C4A and C4B genes. *J. Immunol.* 162:3687–3693.
 28. Porter, R.R. 1983. Complement polymorphism, the major histocompatibility complex and associated diseases: a speculation. *Mol. Biol. Med.* 1:161–168.
 29. Fielder, A.H.L., M.J. Walport, J.R. Batchelor, R.I. Rynes, C.M. Black, I.A. Dodi, and G.R.V. Hughes. 1983. Family study of the major histocompatibility complex in patients with systemic lupus erythematosus: importance of null alleles of C4A and C4B in determining disease susceptibility. *Br. Med. J.* 286:425–428.
 30. Christiansen, F.T., R.L. Dawkins, and G. Uko. 1983. Complement allotyping in SLE: association with C4A null. *Aust. NZ J. Med.* 13:483–488.
 31. Atkinson, J.P., and P.M. Schneider. 1999. Genetic susceptibility and class III complement genes. In *Systemic Lupus Erythematosus*. R.G. Lahita, editor. Academic Press, San Diego. 91–104.
 32. New, M.I., and R.C. Wilson. 1999. Steroid disorders in children: congenital adrenal hyperplasia and apparent mineralocorticoid excess. *Proc. Natl. Acad. Sci. USA.* 96:12790–12797.
 33. Rupert, K.L., R.M. Rennebohm, and C.Y. Yu. 1999. An unequal crossover between the RCCX modules of the human MHC leading to the presence of a CYP21B gene and a tenascin TNXB/TNXA-RP2 recombinant between C4A and C4B genes in a patient with juvenile rheumatoid arthritis. *Exp. Clin. Immunogenet.* 16:81–97.
 34. Sim, E., and S. Cross. 1986. Phenotyping of human complement component C4, a class III HLA antigen. *Biochem. J.* 239:763–767.
 35. Awdeh, Z.L., and C.A. Alper. 1980. Inherited structural polymorphism of the fourth component of human complement. *Proc. Natl. Acad. Sci. USA.* 77:3576–3580.
 36. Schneider, P.M., and C. Rittner. 1997. Complement genetics. In *Complement: A Practical Approach*. A.W. Dodds and R.B. Sim, editors. IRL Press at Oxford University Press, Oxford. 165–198.
 37. O'Neill, G.J., S.Y. Yang, and B. DuPont. 1978. Two HLA-linked loci controlling the fourth component of human complement. *Proc. Natl. Acad. Sci. USA.* 75:5165–5169.
 38. Roos, M.H., E. Mollenhauer, P. Demant, and C. Rittner. 1982. A molecular basis for the two locus model of human complement component C4. *Nature.* 298:854–856.
 39. Carroll, M.C., T. Belt, A. Palsdottir, and R.R. Porter. 1984. Structure and organization of the C4 genes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 306:379–388.
 40. Rebmann, V., I. Doxiadis, B.S. Kubens, and H. Grosse-Wilde. 1992. Quantitation of the human component C4: definition of C4 Q0 alleles and C4A duplications. *Vox Sang.* 62:117–123.
 41. Holme, E., S.J. Cross, J. Veitch, G.J. O'Neil, and K. Whaley. 1988. Quantitation of human C4A and C4B, in serum and plasma by enzyme-linked immunosorbent assay. *Immunogenetics.* 27:295–297.
 42. Moulds, J.M., F.C. Arnett, C.G. Giles, and R.G. Hamilton. 1990. A novel immunoassay for the quantitation of human C4 gene products. *Complement Inflamm.* 7:95–101.
 43. Higashi, Y., H. Yoshioka, M. Yamane, O. Gotoh, and Y. Fujii-Kuriyama. 1986. Complete nucleotide sequence of two steroid 21-hydroxylase genes tandemly arranged in human chromosome: a pseudogene and a genuine gene. *Proc. Natl. Acad. Sci. USA.* 83:2841–2845.

44. Rodrigues, N.R., I. Dunham, C.Y. Yu, M.C. Carroll, R.R. Porter, and R.D. Campbell. 1987. Molecular characterization of the HLA-linked steroid 21-hydroxylase B gene from an individual with congenital adrenal hyperplasia. *EMBO (Eur. Mol. Biol. Organ.) J.* 6:1653-1661.
45. Witchel, S.F. 1999. Molecular diagnosis in the congenital adrenal hyperplasias. *The Endocrinologist.* 9:197-207.
46. White, P.C., A. Vitek, B. DuPont, and M.I. New. 1988. Characterization of frequent deletions causing steroid 21-hydroxylase deficiency. *Proc. Natl. Acad. Sci. USA.* 85:4436-4440.
47. Awdeh, Z.L., D. Raum, E.J. Yunis, and C.A. Alper. 1983. Extended HLA/complement allele haplotypes: evidence for T/t-like complex in man. *Proc. Natl. Acad. Sci. USA.* 80: 259-263.
48. Simon, S., L. Truedsson, D. Marcus-Bagley, Z. Awdeh, G.S. Eisenbarth, S.J. Brink, E.J. Yunis, and C.A. Alper. 1997. Relationship between protein complotypes and DNA variant haplotypes: complotypes-RFLP constellations (CRC). *Hum. Immunol.* 57:27-36.
49. Gatenby, P.A., J.E. Barbosa, and P.J. Lachmann. 1990. Differences between C4A and C4B in the handling of immune complexes: the enhancement of CR1 binding is more important than the inhibition of immunoprecipitation. *Clin. Exp. Immunol.* 79:158-163.