THE GENETIC COMPLEXITY AND PROTEIN POLYMORPHISM OF COMPLEMENT C4 IN HEALTH AND DISEASE

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By

Yan Yang, B.M.

* * * * *

The Ohio State University
2004

Dissertation Committee:

Professor Chack-Yung Yu, Adviser
Professor Caroline C. Whitacre
Professor Marshall Vance Williams Jr
Professor Lee Alois Hebert

Approved by

__________________________
Adviser
Department of Molecular Virology, Immunology and Medical Genetics
Human complement C4, located in Major Histocompatibility Complex class III region, is an essential component of both the classical and the lectin activation pathways of the complement system. The protein products are categorized into two isotypes: C4A and C4B that may exhibit different biological functions. Both C4A and C4B have unusually frequent variations in the gene number and a wide range of serum protein levels among different individuals and ethnic groups (100 -1000 µg/ml). The number of C4 genes present in an individual varies from 2 to 6. The variation of C4 gene dosage is always concurrent with three other flanking genes, including RP, CYP21 and TNX. The size of C4 genes can be long (21 kb) and short (14.6 kb), which differs in 6.4 kb due to the presence or absence of an endogenous retroviral sequence in intron 9. Complete or partial deficiency of C4A has been observed to be a risk factor for systemic lupus erythematosus (SLE). However, the deeply ingrained “two-locus” model of human C4 genes greatly hindered its epidemiological study in humans.

The first objective of my dissertation was to determine the genetic complexity and protein polymorphism of C4 in Caucasians. It was observed that the protein levels of C4, C4A and C4B are determined by the corresponding gene dosages, gene size and body mass index. There were good correlations between the number of long C4 genes with C4A protein levels, and the number of short C4 genes with C4B protein levels. The short
C4B gene in the second locus of LS RCCX haplotype not only increased the C4B protein level, but also increased the C4A protein level that was expressed by the long C4A gene in the first locus.

The second objective of my dissertation was to determine the molecular basis of C4 mutations from complete C4 deficient subjects. Three novel C4 mutations were discovered in eight C4 deficient individuals from US and Austria, which included a single C-deletion in exon 13 of a short C4B gene, a GT-deletion in exon 13 of a long C4A gene, and a G to A mutation at the 5’ donor site of intron 28 in two short C4B genes. SS-PCR strategies were created to allow a rapid screening of known C4 mutations in human populations.

The third objective of this dissertation was to study variations of C4 gene dosages and protein polymorphisms in Caucasian and Black SLE patients. When compared with case controls, it was shown that deficiencies of total C4 and C4A genes were increased in both Caucasian and Black SLE patients. Intriguingly, protein levels of total C4, C4A and C4B decreased in SLE patients, even within the same gene dosage group. The low protein levels of total C4 and C4A in SLE were caused by deficiencies of C4 and C4A genes, C4A mutations, and low C4 expression or high C4 consumption, which may in turn lead to inefficient clearance of immune complexes (ICs), deposition of ICs in local tissues, and subsequent tissue injury. The study in my dissertation provides valuable information on C4 polygenic variations in humans that may help diagnosis and therapeutic intervention of C4-related diseases, especially SLE.
Dedicated to my parents, husband and daughter
ACKNOWLEDGMENTS

First of all, I wish to give my sincere thanks to my advisor, Dr. Chack-Yung Yu, for his intellectual support, encouragement, and enthusiasm. As my best friend and an experienced mentor, he is always willing to provide me with his support and help. It is with his trust that I can conquer the mountain eventually. I am also very grateful for the opportunities that he has provided for me to think and solve scientific problems independently and to present my findings in the audience of peers in national meetings.

I wish to thank my dear fellow student, Erwin Chung, for his unselfish help and intellectual inspiration, especially in the first year when I came to the US. Although he left our lab last year after getting the Ph.D. degree, his generosity has made us very good friends even now. I also appreciate him very much for his contributions for the SLE project.

My appreciation goes to Bi Zhou for the beautiful experiments she has done in the last several years. I thank her for her time spent on the SLE project and normal populations. To my current and past colleagues who work in our laboratory, Yao-ling Shu, Nora Alghothani, Chris Bailey, Denise Yu, Natalie Jacobsen, Dr. Kapil Saxena, Yee lin Wu, and Jenny Kohout, thanks for your help and companionship.
I wish to express my gratitude to Dr. Lee Hebert for his encouragement and guidance during my Ph.D. study. I also wish to thank him and his wife, Maddie Hebert for their great contributions to the Ohio SLE study.

I wish to express my appreciation to Dr. Marshall Williams for his encouragements to my Ph.D. study. I am indebted to Dr. Caroline Whitacre, who admitted me into our Ph.D. program and watched me walk through the course of my Ph.D. study. As my graduate committee member, she has been very supportive of my academic pursuits at OSU.

I wish to thank Ms Charlene Cameron for her help in our publication submission. I am grateful for her warm support and enthusiasm in all these years.

Finally, I wish to thank my parents for their spiritual support in all these years. Although I had left them over four years, they always provide me with the love, time and effort that they can possibly do. I am grateful for my husband, Bin, for the support, encouragement and accompany that he provided since he met me, especially after we left home to this foreign country. My gratitude also goes to my parent-in-laws, for their help and support.
VITA

December 26, 1973..........Born in Chang-sha, Hunan, China

1992-1997.........................Bachelor of Medicine, Institute of Clinical Medicine,  
Xiang-Ya School of Medicine, Central South University,  
Chang-Sha, Hunan, China

1999-2004.........................Graduate Teaching and Research Associate  
Department of Molecular Virology, Immunology, and  
Medical Genetics, The Ohio State University, Columbus,  
Ohio, United States of America

PUBLICATIONS

1. Blanchong CA, Chung EK, Rupert KL, Yang Y, Yang Z, Zhou B, Moulds JM, and  
Yu CY: Genetic, structural and functional diversities of human complement  
components C4A and C4B and their mouse homologues, Slp and C4. Internat  

Dawkins RL, and Yu CY: Genetics of human complement component C4 and  

3. Yu CY, Blanchong CA, Chung EK, Rupert KL, Yang Y, Yang Z, Zhou B and  
Moulds JM: Molecular genetic analyses of human complement components C4A and  
C4B. Manual of Clinical Laboratory Immunology, 6th Edition, American Society for  

4. Rupert KL, Moulds JM, Yang Y, Arnett FC, Warren RW, Reveille JD, Myones BL,  
Blanchong CA and Yu CY: The molecular basis of complete complement C4A and  
C4B deficiencies in a systemic lupus erythematosus (SLE) patient with homozygous  

5. Chung EK, Yang Y, Rupert KL, Jones KN, Rennebohm RM, Blanchong CA and Yu  
CY: Determining the 1, 2, 3 or 4 long and short loci of human complement C4 in  
major histocompatibility complex haplotype coding for C4A or C4B proteins. Am J  


**FIELDS OF STUDY**

Major Field: Molecular Virology, Immunology and Medical Genetics
# TABLE OF CONTENTS

Dedication.................................................................................................................iv

Acknowledgements......................................................................................................v

Vita..........................................................................................................................vii

List of Tables.............................................................................................................xiv

List of Figures...........................................................................................................xvi

Chapters:

1. Introduction.............................................................................................................1

2. Diversity in intrinsic strengths of the human complement system: serum C4 protein concentrations correlate with C4 gene size and polygenic variations, hemolytic activities, and body mass index.................................................................62

   2.1 Abstract.............................................................................................................62

   2.2 Introduction......................................................................................................64

   2.3 Materials and methods..................................................................................67

      2.3.1 Study population.......................................................................................67

      2.3.2 Measurement of serum C4 and C3 concentrations.................................67

      2.3.3 Measurement of the complement C4HA..................................................67

      2.3.4 Preparation of genomic DNA, Southern blot analysis, and DNA probes.................................................................68
2.3.5 Determination of C4 protein polymorphism by immunfixation and immunoblot analysis.................................................................69
2.3.6 Statistical analysis........................................................................69
2.3.7 Additional declarations.................................................................69
2.4 Results...............................................................................................71
  2.4.1 Genotypic variation of human complement components C4A and C4B.................................................................................71
  2.4.2 Qualitative variations of C4A and C4B........................................74
  2.4.3 Quantitative diversities of C4A and C4B proteins.........................75
  2.4.4 Hemolytic activities, C4 gene dosages, C4 gene sizes and serum C4 concentrations.................................................................77
  2.4.5 BMI and serum C4 and C3 concentrations....................................78
2.5 Discussion.........................................................................................81
  3.1 Abstract..........................................................................................98
  3.2 Introduction....................................................................................100
  3.3 Materials and methods.................................................................103
    3.3.1 Human subjects and peripheral blood samples......................103
    3.3.2 Allotyping of human complement components C4A and C4B proteins.................................................................103
    3.3.3 Genomic RFLPs.....................................................................103
    3.3.4 Specific oligonucleotide primers used in this study..............104

x
3.3.5 PCR amplification, cloning and sequencing of long C4A mutant gene from HLA A24 B38 DR13 ......................................................... 105

3.3.6 SSP-PCR (Sequence-Specific Primer PCR) for the 2-bp deletion at exon 13 ................................................................. 106

3.3.7 PCR amplification, cloning and sequencing of two short C4B mutant genes from HLA A30 B18 DR7 ............................................ 106

3.3.8 SSP-PCR-MboI RFLP to detect mutation at the 5’ splice site of intron 28 ................................................................. 107

3.3.9 PCR for exon 9-11 to segregate two mutant C4B genes .............. 107

3.3.10 Additional declarations ...................................................... 108

3.4 Results .................................................................................. 109

3.4.1 Patients’ clinical histories ...................................................... 109

3.4.2 RCCX modules, C4A and C4B mutant genes in the seven complete C4 deficiency patients ........................................ 113

3.5 Discussions .......................................................................... 120

4. The association study of complement C4, C2 and TNFA with systemic lupus erythematosus (SLE) ......................................................... 133

4.1 Abstract ................................................................................ 133

4.2 Introduction .......................................................................... 135

4.3 Materials and methods .............................................................. 141

4.3.1 Recruitment of human subjects and preparation of DNA and EDTA-plasma .............................................................. 141
4.3.2 Determination of the number and size of C4A and C4B genes and RCCX structures by TaqI and PshAl-PvuII RFLPs (Restriction Fragment Length Polymorphism)……………………………………141

4.3.3 Ascertainment of the RCCX haplotype and number and size of C4 genes on each chromosome 6 by PmeI PFGE (Pulsed Field Gel Electrophoresis)……………………………………………………142

4.3.4 Determination of C4A and C4B allotypes by C4 immunofixation….143

4.3.5 Determination of total C4 levels in plasma by radial immunodiffusion (RID)………………………………………………………………143

4.3.6 Determination of C4A and C4B concentrations…………………………143

4.3.7 Detection of plasma C4 split product C4a levels by ELISA……..143

4.3.8 Detection of known C4 and C2 mutations in a single multiplex PCR reaction…………………………………………………………………….144

4.3.9 Determination of TNFA -308 polymorphism by SSP-PCR (Sequence Specific-Primer-Polymerase Chain Reaction)…………………………144

4.3.10 Statistical analyses…………………………………………………….145

4.3.11 Additional declarations……………………………………………….145

4.4 Results………………………………………………………………………...146

4.4.1 The differences in C4 polygenic and RCCX modular variations between female healthy Caucasians and Caucasian SLE families………………146

4.4.2 The variation in total C4, C4A and C4B protein levels in female Caucasian SLE patients and unaffected family members………………148

4.4.3 C4 mutations in Caucasian SLE…………………………………………149
4.4.4 The effect of sex on C4 genetic variation and C4 protein expression in
SLE patients and unaffected family members……………………………..150
4.4.5 The effect of other susceptibility genes in MHC region on Caucasian
SLE……………………………………………………………………150
4.4.6 The comparison of C4 polygenic and RCCX modular variations and C4
protein expression between normal African Americans and African
American SLE patients…………………………………………………152
4.4.7 The complement C4 activation in SLE patients……………………153
4.4.8 A comparison of C4 polygenic and RCCX modular variations and C4
protein expression between Caucasian and African American SLE
patients………………………………………………………………154
4.5 Discussions………………………………………………………………….156
5. General Discussion………………………………………………………………175

Bibliography……………………………………………………………………185
LIST OF TABLES

Table  Page

2.1  A summary of RCCX length variants and C4 gene dosages in Caucasians........94

2.2  Phenotypes of complement components C4A and C4B in a healthy Hungarian population..........................................................96

2.3  Linear regression analysis of the relationship between BMI and serum C3 and C4 concentrations in healthy Hungarian subjects...........................................97

3.1  A summary of clinical histories in 7 complete C4 deficiency patients..........130

3.2  The sequence changes in the long C4A genes in families 2 and 4.................131

3.3  Nucleotide changes in the short C4B genes from patient 1P......................132

4.1  The variations in C4, C4A and C4B gene dosages and RCCX modular structures among female normal Caucasians, female Caucasian SLE patients, mothers and siblings.................................................................170

4.2  The differences in the variations of C4, C4A and C4B gene dosages and RCCX modular structures between female and male Caucasian SLE patients...........171

4.3  The polymorphism of TNFA at –308 in Caucasian SLE and unrelated normal controls..............................................................172
4.4 A comparison of the variations in total C4, C4A and C4B gene dosages and RCCX modules between normal African Americans and African American SLE patients…………………………………………………………………………...173

4.5 A comparison of the variations in total C4, C4A and C4B gene dosages and RCCX modules between Caucasian and African American SLE patients…………………174
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>The three activation pathways of complement</td>
</tr>
<tr>
<td>1.2</td>
<td>A molecular map of the genes in the MHC complement gene cluster</td>
</tr>
<tr>
<td>1.3</td>
<td>The polygenic and gene size variations of human complement C4 and RP-C4-CYP21-TNX (RCCX modules) in the MHC class III region</td>
</tr>
<tr>
<td>1.4</td>
<td>The qualitative variations of complement C4A and C4B proteins</td>
</tr>
<tr>
<td>1.5</td>
<td>Protein structure of human C4</td>
</tr>
<tr>
<td>1.6</td>
<td>The sophisticated genetic diversity of human C4 genes</td>
</tr>
<tr>
<td>1.7</td>
<td>A comparison of frequencies of homozygous and heterozygous C4AQ0 in SLE patients and matched controls in various ethnic groups/races</td>
</tr>
<tr>
<td>2.1</td>
<td>The gene size and polygenic variations of human complement C4 and RP-C4-CYP21-TNX (RCCX modules) in the MHC class III region</td>
</tr>
<tr>
<td>2.2</td>
<td>Genotypic variations of RCCX modules, C4A and C4B gene dosages, and protein polymorphisms</td>
</tr>
<tr>
<td>2.3</td>
<td>Analyses of C4 serum concentrations with C4 gene dosage and gene size variations</td>
</tr>
</tbody>
</table>
2.4 Analyses of C4 hemolytic activities with C4 polygenic and gene size variations, and with serum C4 concentrations...............................................................91

2.5 The impact of long and short C4 genes on the serum C4 protein concentrations and C4HA...........................................................................................................92

2.6 Analyses of BMI and C4 polygenic variations, serum C4 concentrations, C4HA, and serum C3 concentration.................................................................93

3.1 Phenotypic and genotypic analyses of complement C4 in seven complete C4 deficiency patients and available family members.................................124

3.2 Determination of CYP21A and CYP21B genes in complete C4 deficient subjects and family members from families 1,3 and 4 using BsaI RFLP..............125

3.3 Monomodular RCCX structure, PCR amplifications and sequence determination of the mutation in the long C4A gene from HLA A24 Cw7 B38 DR13...........126

3.4 Bimodular SS RCCX structures, PCR amplification and sequence determination of the mutations in the two short C4B genes from HLA A30 B18 DR7............127

3.5 SSP-PCRs to determine the novel mutations in the long C4A gene and the two short C4B genes, and the sequence polymorphisms between these two short C4B genes.................................................................................................................128

3.6 The hotspots of deleterious mutations in C4AQ0 and C4BQ0 in humans........129

4.1 A comparison of the variations in gene dosages of total C4, C4A, C4B and RCCX structures between normal and SLE Caucasians..................................................161

4.2 A comparison of total C4, C4A and C4B protein levels among female Caucasian SLE patients, mothers and siblings.................................................................162
4.3 Correlations of total C4, C4A and C4B protein levels with the corresponding gene
dosages in female SLE patients, mothers and female siblings......................163

4.4 A comparison of the mean total C4, C4A and C4B protein levels between female
and male Caucasian SLE patients.................................................................164

4.5 A comparison of the mean total C4, C4A and C4B protein levels among female and
male Caucasian SLE family members..........................................................165

4.6 A comparison of the total C4, C4A and C4B protein levels among normal African
Americans and African American SLE patients.........................................166

4.7 A comparison of the total C4, C4A and C4B protein levels among African
American and Caucasian SLE patients.........................................................167

4.8 A comparison of complement C4 activation product C4a in the circulation between
African American and Caucasian SLE patients..............................................168

4.9 The effect of total C4, C4A and C4B protein levels on the plasma levels of
complement C4 activation product C4a.........................................................169
CHAPTER 1

INTRODUCTION

The immune system includes a group of cells and molecules with specialized roles in defending against infection. There are two fundamentally different types of immunities to invading microbes, that is, innate and adaptive immunities. The innate immunity includes phagocytic cells, cells that release inflammatory mediators, natural killer cells and molecules such as complement, acute-phase proteins, and cytokines (Delves and Roitt 2000). The innate immunity, which is highly conserved even in the simplest animals, confirms its importance in survival. Adaptive immunity, present only in vertebrates, is the hallmark of the immune system. It consists of antigen-specific reactions through T and B lymphocytes. B cells secrete immunoglobulins, the antigen-specific antibodies responsible for eliminating extracellular microorganisms. T cells help B cells to make antibody and can also eradicate intracellular pathogens by activating macrophages and by killing virally infected cells. Innate and adaptive immunity usually work together to eliminate pathogens. The essential function of the immune system in host defense is best demonstrated when it goes wrong; underactivity resulting in the severe bacterial and viral infections, overactivity in allegic and autoimmune diseases (Delves and Roitt 2000, Parkin and Cohen 2001).
Although complement is part of innate immune response involved in direct killing of invaded microbes, it is suggested that complement can bridge between innate and adaptive immunity by augmenting antibody response and enhancing immunologic memory. In addition, complement may play a role in the clearance of immune complex and apoptotic bodies, which make it a candidate for autoimmune disease susceptibility gene. The underlying pathogenesis of autoimmune disease is the production of autoantibodies against autoantigens such as those found in apoptotic bodies and/or the activation of auto-reactive T cells.

Complement C4, with its variations in gene number, size and nucleotide polymorphisms, is particularly relevant. Studies of complete C4 deficiency in human and mouse all suggest its role in autoimmunity, especially, systemic lupus erythematosus (SLE). In this chapter, the complex genetics of complement C4 and the current opinion of its association with autoimmune diseases, including SLE are reviewed.

1.1 The Complement System

Complement was first identified as a heat-labile principle in serum that “complemented” antibodies in the killing of bacteria (Walport 2001b). The term “complement” was introduced by Paul Ehrlich and Julius Morgenroth (Klein 1982). Following the realization that complement combines with antibodies to kill not only bacteria but also eukaryotic cells, Bordet simplified the study of complement by designing an assay system consisting of sheep erythrocytes as target cells, rabbit antibodies to sheep red cells, and guinea pig serum as the source of complement. The assay, called hemolytic activity, is used to this day; its main advantage is that
complement activity can be measured easily through the release of hemoglobin from the lysed erythrocytes. In the circulation, complement proteins amount to more than 3 g per liter and constitute approximately 15 percent of the globulin fraction. Complement proteins on cell membranes can be receptors for activated complement proteins or proteins that regulate complement activity. In total, more than 30 complement proteins in plasma or on cell surfaces have been identified (Walport 2001b, Walport 2001a).

Complement system is part of innate immunity. Complement proteins can be activated in a cascade sequence, with amplification stages. There are three pathways of complement activation that can be driven by the presence of a foreign substance: the classical pathway by antigen-antibody reactions, the alternative pathway by polysaccharides from yeasts and gram negative bacteria, and the lectin pathway by mannose containing proteins and carbohydrates on microbes, including viruses and yeasts. All three pathways converge with the activation of complement C3, leading to the formation of membrane attack complex on the cell surface and osmotic cell lysis (Figure 1.1). Complement activation is usually focused on the surface of the cells from invading microorganisms, where the complement inhibitory proteins have limited access. Normal host cells bear membrane cofactor protein (MCP), the complement receptor 1 (CR1), and/or decay accelerating factor (DAF), which can inhibit C3 convertase and prevent progression of complement activation (Parkin and Cohen 2001).

Complement has three overarching physiologic activities: defending against pyogenic bacterial infection, bridging innate and adaptive immunity, and disposing of immune complexes and the products of inflammatory response (Walport 2001b). Deficiencies in complement proteins, such as C3, C7, C9 and MBL, can increase
susceptibility to bacterial and viral infections (Lachmann 1987, Ross and Densen 1984, Super et al 1989, Wurzner et al 1992). The role of complement in the induction of antibody responses was first demonstrated by Pepys that the formation of antibodies against T-cell-dependent antigens was reduced in animals in which C3 had been depleted (Pepys 1974). It was suggested that binding of C3 split product C3d to the complement receptor 2 (CR2) on B cell surface may help lower the threshold for B cell activation (Fearon and Carroll M 2000). Under physiological conditions, complement promotes the clearance of immune complexes, an important means of eliminating antibody-coated bacteria. If, however, immune complexes cannot be eliminated, then complement becomes chronically activated and can incite inflammation. In autoimmune diseases such as SLE, the activation of complement by immune complexes, comprised of autoantibodies and autoantigens that cannot be cleared from the body effectively, is an important contributor to tissue injuries.

1.2 Complement C4

Complement C4 is an essential component of both classical and lectin complement activation pathways. The activated C4 is a subunit of the C3 and C5 convertases, the enzymatic complexes that activate C3 and C5 leading to the formation of membrane attack complex on cell surface and subsequent cell lysis. Complement C4 was discovered by J. Gordon and his colleagues in 1926, as its activity in serum was shown to withstand heating to 56 degrees, was not affected by mixing with yeast cells, but could be inactivated by treatment with ammonia or hydrazine (Gordon et al 1926, Klein 1982). Evolutionally, C4 emerged in the vertebrates as part of a big bang (Du Pasquier and
Litman 2000, Rached et al 1999). The remarkably frequent variation in gene dosage and
gene size, diversities in the protein primary structures, expression levels, and functions of
human C4A and C4B are probably some of the mechanisms developed by the vertebral
immune system to overcome a vast variety of microbial infections. C4 deficiencies may
lead to defective processing of immune complexes, impairment of B-cell memory, and/or
persistence of bacterial and viral infections (Carroll 1998b, Colten and Rosen 1992,
Hauptmann et al 1988). On the other hand, excessive production of complement C4
could potentially cause over-activation of the complement pathways and exacerbate the
inflammatory response at the local tissues.

1.2.1. Genetic diversity of human complement C4A and C4B

The human C4 genes are located in the Major Histocompatibility Complex (MHC)
class III region of chromosome 6 (Figure 1.2). The variations of C4 are shown in gene
number, size and nucleotide polymorphisms.

In humans, the number of C4 genes present in an individual varies from 2 to 6
(Blanchong et al 2000)(Figure 1.3). The variation of C4 gene dosage is always concurrent
with three other flanking genes, including the serine/threonine nuclear protein kinase RP
(also termed as STK19), steroid 21-hydroxylase CYP21 and extracellular matrix protein
tenascin TNX in the MHC complement gene cluster (Yu 1998, Yu et al 2000). This RP-
C4-CYP21-TNX (RCCX) modular variation is frequent in all populations (Blanchong et
et al 1985, Yang et al 1999). When duplicated into a bimodular structure, an additional
functional gene is generated for complement component C4, but the other three
concurrently duplicated constituents located between the C4 genes are usually nonfunctional.

The size of C4 genes can be long (21 kb) and short (14.6 kb) due to the integration of the endogenous retrovirus HERV-K(C4) in intron 9 of the long gene (Belt et al 1985, Carroll et al 1985, Dangel et al 1994, Heidmann et al 1988, Schneider et al 1986, Simon et al 1997, Yu and Campbell 1987). The size of intron 9 is 6782-bp in a long C4 gene, but only 416-bp in a short C4 gene. The site of integration is 283-bp downstream of the intron donor site. The configuration of this endogenous retrovirus is opposite to the transcriptional orientation of C4.

In humans, C4 can be classified into two isotypes, C4A and C4B, depending on 5 nucleotide differences at positions 7609, 7612, 7620, 7623 and 7625 in the C4d region of the α-chain (Yu et al 1986). This region also determines the Rodgers (Rg) versus Chido (Ch) antigenic determinants (Tilley et al 1978). C4A usually associates with Rg blood group antigens, while C4B associates with Ch blood group antigens (Giles et al 1988, Moulds et al 1996, Yu et al 1988). Altogether, over 41 C4 protein allotypes, determined by both common and specific nucleotide polymorphisms that lead to charge and serological differences in the plasma C4 proteins, have been identified (Mauff et al 1998)(Figure 1.4).

1.2.2 Endogenous retrovirus in long C4 gene

The human complement C4 genes in the HLA exhibit an unusual, dichotomous size polymorphism mediated by the endogenous retrovirus, HERV-K(C4). The name HERV-K(C4) is derived from a characteristic primer binding site (PBS) for the transfer RNA for lysine (i.e., K). Nearly identical sequences for this retrotransposon are present at
precisely the same location in the long C4 genes in reverse orientation (Dangel et al 1994). Mutations accumulated through evolution make it highly unlikely to produce a live retrovirus, although retrovirus-related transcripts or protein products could be generated (Dangel et al 1994). The sequences for the 5' LTR and the 3' LTR share 89% sequence identities, including a TATA-box and an SV40-type transcriptional enhancer. *In vitro* transient expression assays have demonstrated the 5' LTR of HERV-K(C4) has lost its promoter activity in both orientations, while the 3' LTR has lost the promoter activity in its sense but not the antisense configuration (Dangel et al 1994).

Evolutionarily, both long and short C4 genes are present in most Old World primates studied. However, only short C4 genes are detected in the New World primate cotton-top tamarin, and in old world primates chimpanzee and gorilla (Dangel et al 1995). A comparison of the short intron 9 sequences from seven different human C4 genes and six other short C4 genes from non-human primates did not reveal any traces of LTR-LTR recombination or solitary LTR sequences flanked with target site repeats, suggesting short C4 genes are the ancestral forms that coexist with long C4 genes. Molecular genetic analysis of >300 healthy Caucasians revealed that HERV-K(C4) is present in 76% of human C4 genes. In other words, about 24% of the human C4 genes are short. Such a size dichotomy poses a question about the evolutionary advantage of maintaining both long and short C4 genes in primate evolution. By using reporter gene assay (Dangel et al 1994, Schneider et al 2001), Schneider and colleagues demonstrated that HERV-K(C4) antisense mRNA transcripts were present in cells constitutively expressing C4, and expression of retroviral-like constructs was significantly downregulated in cells expressing C4. This downregulation was further modulated in a dose-dependent fashion
following interferon γ stimulation (Schneider et al 2001). Therefore, expression of C4 could lead to the transcription of an antisense HERV RNA, which might protect against exogenous retroviral infections by forming a double-stranded RNA molecule to trigger the antiviral response through the interferon system. However, the effect of this retroviral element on C4 expression is unknown.

1.2.3 Structural and physiological characteristics of C4A and C4B

The C4A/C4B isotypic residues are PCPVLD 1101-6 LSXPVIH (Figure 1.5). In vitro, the reaction rate of activated C4A was estimated to be more than 20-fold slower than that of C4B. Site-directed mutagenesis indicates that D1106 is the critical residue affecting the enhanced binding of C4A towards amino group containing antigens, while H1106 is responsible for the catalytic binding of C4B to hydroxyl groups on targets (Carroll et al 1990, Dodds et al 1996, Isenman and Young 1984, Law et al 1984). Sources of amino groups for activated C4A include immune complex (IC) and complement receptor 1 (CR1). A source of hydroxyl groups for activated C4B is carbohydrates on bacterial cell walls or cell membranes. Therefore, the in vitro studies suggest that C4A is more important in the clearance of IC, while C4B is important in the propagation of complement activation pathways. The X-ray crystal structure of human C4d with the location of D/H1106 at the convex surface together with the thioester C991 and Q994 further supports the role of H1106 in the catalysis of the thioester carbonyl group from Q994 in a trans-esterification reaction to form covalent ester bonds with hydroxyl groups of substrates (Dodds et al 1996, van den Elsen et al 2002).

Alloantibodies against complement C4A and C4B are formed in blood transfusion recipients who have C4A or C4B protein deficiency, or if there are mismatches in the
variants of C4A or C4B between the donor and the recipient. It is suggested that C4A isotype is usually associated with Rg blood group antigens and C4B isotype is associated with Ch blood group antigens (Figure 1.5) (Giles et al 1988, Moulds et al 1996, Yu et al 1988). The Rg and Ch blood group antigens are formed by the deposition of the activated C4, C4b, on the membranes of erythrocytes. When CR1 and factor I are present, C4b can be further cleaved to yield the soluble C4c and the membrane bound C4d (Atkinson et al 1988, Giles et al 1988, O'Neill et al 1978b, Tilley et al 1978). Two Rg (Rg1 and Rg2) and six Ch (Ch1 to Ch6) epitopes were defined from a battery of alloantibodies against C4. The major determinants Rg1 and Ch1 are defined by VDLL in C4A and ADLR in C4B at residues 1188-91, respectively. The C4B isotypic residues LSPVIH 1101-6 are associated with Ch4. However, the C4A isotypic residues PCPVLD 1101-6 are not associated with Rg antigenic determinants (Yu et al 1988). After transfusion, the generation of such alloantibodies could cause adverse transfusion reactions (Giles and Swanson 1984, Lambin et al 1984).

1.2.4 The C4 activation product C4a

The activation of C4 leads to the proteolysis of N-terminal region of the alpha chain, generating a 77 amino acid peptide with neoepitope, C4a (residues 616-737). Same as C5a and C3a, C4a is an anaphylatoxin (Greer 1986, Hugli 1986). During inflammation, anaphylatoxins function to promote contraction of smooth muscles, increase vascular permeability and platelet aggregation, induce the release of histamine by degranulation of mast cells and stimulate the secretion of interleukin-1 from monocytes. However, it is suggested that the efficiency of C4a as an inflammatory agent is 100 to 1000 foldless than that of C5a. Additional studies on C4a suggest that it might play an anti-
inflammatory role by inducing the release of a 20 kDa protein from monocytes that inhibits its own chemotaxis (Matsubara et al 1991, Tsuruta et al 1993). Using guinea pig macrophages as target cells, human C4a could sensitize macrophages by inducing a biphasic Ca\(^{+2}\) mobilization, but the C4-sensitized macrophages could not react to C4a again (Murakami et al 1993). Moreover, human C4a was unable to compete against human C3a on binding to guinea-pig macrophages (Murakami et al 1993) and had no effects on Ca\(^{+2}\) mobilization in neutrophils or cells stably expressing cloned human C3a receptors, suggesting the cellular receptor for C4a is distinct from C3a receptor (Ames et al 1997, Lienenklaus et al 1998). Future study on identifying receptors for C4a might shed light on its physiologic roles. Besides its own physiological roles, detection of C4a levels in the plasma and/or in the tissue fluid could be used as an indicator for the status of C4 activation during the disease.

1.2.5 Detection of C4A and C4B in human populations

Complement C4 polymorphisms were first demonstrated in 1969 by Rosenfeld, Ruddy and Austen using antibody-antigen crossed electrophoresis (Rosenfeld et al 1969). A single locus with co-dominant expression of fast- and slow-migrating forms of C4 was proposed (Teisberg et al 1976) but soon replaced by a 2-locus model with one C4A and one C4B gene (O'Neill et al 1978a). In this model, differential plasma levels of C4A and C4B proteins were explained by the presence of null or silent alleles in either the C4A or the C4B locus (Awdeh et al 1979, Roos et al 1982). However, homoexpression of C4A proteins from both C4 loci and haplotypes with three loci coding for different C4A and C4B allotypes (Carroll et al 1984, Olaisen et al 1980, Raum et al 1984, Rittner et al 1984, Schendel et al 1985, Teisberg et al 1976, Teisberg et al 1988) were also observed, which

*Taq*I and *Psh*AI-*Pvu*II RFLPs give excellent account of the number of long and short C4A and C4B genes present in a diploid genome. Results for the number and size of C4 genes and RCCX structures can be obtained rapidly from the *Taq*I RFLP, as three different probes specific for RP-C4, CYP21 and TNX are applied to the hybridization mixture simultaneously. *Psh*AI, recognizing the isotypic site of C4A and C4B, allows the differentiation of C4A and C4B genes. The addition of another restriction enzyme, *Pvu*II, helps further reducing the *Psh*AI-digested fragment of both C4A and C4B, allowing a clear assessment of the relative dosages of C4A and C4B genes. *Psh*AI-*Pvu*II RFLP also offers an additional method to confirm RCCX modules deduced by the *Taq*I RFLP. In order to determine the RCCX modules and number and size of C4 genes per chromosome 6, a long-range mapping method, employing pulsed field gel electrophoresis of high molecular weight genomic DNA digested by *Pme*I restriction enzyme, is applied. The *Pme*I-digested fragments are predictable, homogeneous, and independent of the tissue origins of the genomic DNA that may have different methylation patterns, therefore reflects the actual size of the DNA fragments with the recognition sequences (Chung et al 2002c) (Figure 1.6).
When the quantity of genomic DNA is limited, two PCR-based methods are developed to facilitate the genotyping. The module-specific PCR yields information on the total number of RCCX modules and therefore the total number of C4 genes present in a diploid genome. A refined "hot-stop" PCR (Uejima et al 2000) coupled with PshAI or XcmI RFLP allows elucidation of the ratio of C4A and C4B genes, and the association of C4 genes with Rg1 or with Ch1 antigenic determinants, respectively (Chung et al 2002c).

C4 immunofixation can be used to determine the allotypes of C4A and C4B proteins based on gross charge difference using goat anti-human C4 antisera (Yu et al 2002).

Through analysis of 150 normal female Caucasians, our laboratory demonstrated that the number of C4 genes present in a diploid genome varied from 2 to 6. About 52% individuals had 4 C4 genes. The frequencies for C4 gene dosage of 2, 3, 5 or 6 in Caucasians were 2%, 25.3%, 17.3% and 3.3%, respectively. Although the 2-locus model with a C4A gene and a C4B gene in the HLA is a deeply ingrained concept in the genetics of human C4, it is important to note that such C4A-C4B configuration only accounts for 55% of the RCCX haplotypes in the Caucasian population (Blanchong et al 2000, Blanchong et al 2001).

1.2.6 Regulation of C4 expression in human; feedback regulation of C4 expression in guinea pigs

In human C4, the major start site for transcription is 51 nucleotides upstream of the sequence for the translation initiation codon (Belt 1984, Belt et al 1985, Yu 1991). No canonical TATA box is present in the 5’ regulatory sequences of C4 genes. By means of reporter gene assays using human HepG2 cells, a biphasic promoter with
proximal and distal positive regulatory regions separated by a negative regulatory region was shown within the 1 kb sequence upstream of the CAP site (Ulgiati et al 2000, Wu 1987). The proximal regulatory region lies within -178 and the CAP site contains an Sp1 site and three E-boxes (Vaishnaw et al 1998). The 3’ E-box between –78 to –73 has been shown responsible for the IFN-γ induction of C4 gene expression in U937 cells (Ulgiati et al 2000).

In addition to SNPs present in the upstream regulatory sequences, two physical entities could have an impact on the expression of human C4A and C4B include: 1) the different genetic environment beyond position –1524 of the C4 genes, it is RP1 specific sequences in the first RCCX module, but is TNXA specific sequences in the second and third RCCX module; 2) the endogenous retrovirus HERV-K(C4) in intron 9 that is located 2,585 nt downstream of the C4 CAP site, which may have positive or negative regulatory motifs that could affect C4 gene expression. In the mouse MHC, the retroviral sequence in the long terminal repeats of the IMP present in intron 3 of the RP1 gene confers the male specific expression of the S1p protein in certain mouse strains (Shreffler et al 1984, Stavenhagen and Robins 1988, Yang and Yu 2000).

Previous expression studies established that the fraction of guinea pig peritoneal macrophages secreting hemolytically active C4 was inversely proportional to the concentration of fluid phase C4 (Auerbach et al 1983). This effect correlated with the binding of native C4 to the macrophage cell surface, and surface C4 binding had no effect on the proportion of macrophages secreting complement C2. It was further demonstrated that fluid phase extracellular C4 specifically inhibited its own production by affecting the transcription or posttranscriptional modification of C4 mRNA, leading to a net decrease
of C4 mRNA without a change in total cellular RNA or RNA specific for factor B (Auerbach et al 1984). However, such a feedback expression mechanism has not been shown in human C4.

1.2.7 Expression and processing of C4A and C4B in liver and extra-hepatic tissues in humans

Plasma C4 is mainly synthesized and secreted by hepatocytes in the liver. It is synthesized as a single-chain precursor molecule ~200 kDa in size and processed to a three-chain (α−β−γ) structure linked by disulfide bonds before secretion (Gigli 1978, Goldberger and Colten 1980, Hall and Colten 1977, Janatova 1986, Schreiber and Muller-Eberhard 1974, Seya et al 1986). Before secretion, C4 undergoes a series of post-translational modifications including sulfation and glycosylation. Early studies on plasma C4 protein indicate the proteolytic processings are incomplete, as 5-7% of the precursor molecule remains unprocessed as single chained pro-C4 or partially processed with two chained structures (β−α γ or β+ α−γ), and 8% of C4 molecules remain as the secretory form (Sim and Cross 1986). The incomplete processing of C4 caused considerable technical difficulty to resolve different C4 allotypes in the plasma by immunofixation because three bands were obtained for each C4A and C4B allotype. An improvement, made by adding neuraminidase and carboxylpeptidase B into the plasma samples to minimize heterogeneity caused by incomplete processing and glycosylations enables different C4A and C4B allotypes to be separated unambiguously based on the gross difference in charge.
Besides hepatic C4 expression, studies have shown that C4 transcripts/proteins are also detected in peripheral blood monocytes, skin fibroblasts, epithelial cells of the intestine, the lung and the kidneys, astrocytes and microglia of the brain, and synovial tissues (Blanchong et al 2001, Colten and Garnier 1998, Morgan and Gasque 1997, Volanakis 1995). The extra-hepatic C4 expression may be important for the local protection and inflammatory response (Colten 1994). A northern blot analysis using human multiple tissue poly(A)^+ RNA blots and a human C4d specific probe was performed in our laboratory. The result demonstrated the highest level of C4 transcripts occurred in the liver, moderate levels in the adrenal cortex, adrenal medulla, thyroid and kidney, and low levels in the heart, ovary, small intestine, thymus, pancreas and spleen (Blanchong et al 2001). During acute phase or inflammatory response, C4 plasma concentration increases 2-3 fold and its synthesis is stimulated by interferon γ in many cell types (Andoh et al 1993, Bristow et al 1993, Kulics et al 1990, Timmerman et al 1995, Tsukamoto et al 1992b, Tsukamoto et al 1992a, Zhou et al 1993).

1.2.8 Diversity of C4 in mouse

As in man, C4 exists in the mouse in two isoforms encoded by two nonallelic genes, tandemly arranged in the S region of the H-2 major histocompatibility complex, one of which, Slp (sex-limited protein), owes its discovery to its peculiar dependence on androgens for expression. C4 and Slp exhibit a variety of genetic variations in activity, structure and regulation (Shreffler et al 1984).

The mouse Slp and C4 genes both have hybrid combination of P, C, I and H at the orthologous positions for C4A/C4B isotypes (Nonaka et al 1984, Nonaka et al 1985, Ogata et al 1989, Ogata and Sepich 1985). They share 95% sequence identities, in
contrast to >99.5% identities between human C4A and C4B. Unlike the clustering of changes at the C4d region in human C4A and C4B, the sequence changes between mouse C4 and Slp scatter throughout the entire molecules except the N-terminal region of the β-chain and the C-terminal region of the γ-chain. Notably, substantial alterations including the EED 736-738 deletion and 5 substitutions are present in a 10 amino acid stretch in Slp that probably abrogate the classical pathway C1s cleavage site.

The levels of mouse plasma C4 and Slp show great variation (Atkinson et al 1982, Parker et al 1980). This may be caused by variation in the gene number or defects in gene expression. The plasma C4 protein level from H-2\textsuperscript{k} is about 20-fold lower than that of H-2w\textsuperscript{7}. It was found that there was an insertion of the repetitive DNA element B2 into intron 13 in the C4 gene of H-2\textsuperscript{k}, which led to aberrant RNA splicing of C4 in liver and lung (Zheng et al 1992), but not in peritoneal macrophages and other organs (Zheng et al 1993). The serum level of Slp in H-2w\textsuperscript{7} is extraordinarily high and is expressed constitutively in both male and female mice (Klein 1975, Shreffler et al 1984). It is found that mouse strains with this haplotype have one regular Slp gene, one regular C4 gene, and three copies of hybrid genes that consists of 1-3 kb of C4 sequence at the 5’ region and Slp sequence at the remaining 3’ region (Levi-Strauss et al 1985a, Levi-Strauss et al 1985b, Miller et al 1992, Pattanakitsakul et al 1990). On the contrary, no Slp protein is made in mouse strains with H-2\textsuperscript{b} haplotype, which is caused by a single G-nucleotide insertion in exon 24 lead to frame shift mutation and truncation of the Slp protein (Rowen et al 1997c, Yang and Yu 2000).

Although Slp has long been considered nonfunctional, it can mediate hemolysis \textit{in vitro} (Beurskens and Dijk 1997). Interestingly, the most common models for SLE-like
phenomena in mice (NZBxNZW, MRL, BxSB, Palmerstone north and C3H/HeJ) involve Slp deficiency (Beurskens and Dijk 1997). In the MRL lpr/lpr mouse, a strain that spontaneously develops SLE, C4 mRNA increased in kidney coincident with the development of nephritis. In addition, a C4 mRNA increase was noted in the lung, heart and intestine and to a lesser extent in liver of lpr/lpr in comparison to the MRL (+/+)
animals (Passwell et al 1988).

1.2.9 C4A and C4B deficiencies and over-expression

Complete deficiency of C4 (both C4A and C4B) is rare but highly deleterious. All but one HLA typed, complete C4 deficient patient suffered SLE, kidney or immune complex disease (Atkinson and Farries 1987). The frequency of a partial C4A or C4B protein deficiency in the normal Caucasian population was previously estimated to be between 25.5%-33.5% (Hauptmann et al 1986). In 150 normal Ohio female Caucasians analyzed by our laboratory, the frequencies of partial C4A and C4B deficiency were 20.1% and 32% respectively.

In mice, complete C4 deficiency causes spontaneous, lupus-like autoimmunity through a mechanism that is independent of CR1/CR2 (Chen et al 2000). C4-deficient mice were generated using gene targeting technology on mouse strain 129 embryonic stem cells with H-2b genetic background that lacks Slp protein expression. After immunization with a T-dependent antigen, C4-deficient mice had reduced number and size of germinal centers in splenic follicles, diminished primary and secondary antibody responses and failure in Ig class-switching from IgM to IgG (Fischer et al 1996). Compared with wild type mice, C4-deficient mice had increased susceptibility to infections and the 50% lethal dose for group B Streptococci infection in C4-deficient
mice was reduced by 25-fold (Wessels et al 1995). Moreover, a bone marrow transplant of C4-deficient mice with bone marrows from C4 wide-type mice could restore the humoral immunity, as it was shown that the transplant mice produced IgG antibodies against immunogens (Gadjeva et al 2002).

Mouse studies offer supports to the essential role of complement C4 in tolerance and autoimmunity (Prodeus et al 1998). In the 129/B6 mixed background, C4-deficient mice had splenomegaly, impairment in clearance of immune complexes and high levels of autoantibodies including antinuclear antibodies and anti-dsDNA with high penetrance in the female mice at 10 months age (Chen et al 2000). The anti-DNA autoreactivity was further shown to penetrate multiple genetic backgrounds including the heterozygous C4+/− or "partial" deficient mice (Paul et al 2002). Comparing with wild type mice, the C4-deficient mice had markedly higher interstitial inflammation in the kidney as shown by increased infiltration of interstitial cells and foci of tubular atrophy (Welch et al 2002). Similar to the cases in humans, only C4 but not C3 was protective against the lupus disease, as shown by experiments using single and double C4 and C3 knockout mice (Einav et al 2002).

Guinea pigs with inherited deficiency of complement C4 have characteristics of immune complex disease (Bottger et al 1986). Consistent with observations in C4-deficient mice, Guinea pigs with inherited C4 deficiency have diminished secondary immune response (Bottger et al 1985, Ellman et al 1971, Ochs et al 1983) and elevated levels of autoantibodies including IgM rheumatic factor (Bitter-Suermann and Burger 1986, Bottger et al 1986, Ellman et al 1970). The predisposition to autoimmune diseases particularly to systemic lupus erythematosus (SLE) is similar to those described in
complete C4A and C4B deficient patients, and homozygous or heterozygous C4A
deficient individuals (Atkinson and Schifferli 1999, Hauptmann et al 1988, Jackson et al
1979).

Crry (complement receptor 1-related protein/gene y) is the complement receptor
type 1 (CR1) in rats. CR1 is a regulator of complement activation that acts by binding to
C3b and C4b to inhibit C3 convertase activity. Rats immunized with anti-Fx1A develop
autoantibodies to Crry as well as to the megalin-containing Heymann nephritis (HN)
antigenic complex. Anti-Crry Abs promote the development of injury in HN by
neutralizing the complement regulatory activity of Crry, which leads to unrestricted
complement activation (Schiller et al 1998). On the other hand, overexpression of Crry
protects mesangial cells from complement mediated injury (Nangaku et al 1997).

1.3 MHC (Major Histocompatibility Complex)

1.3.1 General concepts of MHC

Human MHC (also termed as HLA-Human Leukocyte Antigen), at 6p21.31 is
among the first multi-megabase regions of the human genome to be completely
sequenced (Dawkins et al 1999). With over 200 identified loci in about 4 Mb of DNA,
the MHC is one of the most gene-dense regions in human genome. It has been divided
into three classes, that is, class I, class II and class III. An intriguing feature of the MHC
is the occurrence of particular combinations of alleles, at loci across this 4 Mb region,
more frequently than would be expected based on the frequencies of individual alleles.
This non-random association of alleles gives rise to extended, or ancestral, haplotypes
A second interesting feature of the MHC is the occurrence of extensive duplication, often involving several linked genes or gene fragments.

The two classes, I and II, encode highly polymorphic molecules that present antigens to matched sets of T cell receptors. The class I gene code for the $\alpha$ polypeptide chain of the class I molecule; the $\beta$ chain of the class I molecule is encoded by a gene on chromosome 15, the $\beta_2$-microglobulin gene. There are some 20 class I genes in the HLA region; three of these, HLA-A, B, and C, the so-called classic, or class Ia genes, are the main actors in the immunologic theater. The class II genes code for the $\alpha$ and $\beta$ polypeptide chains of the class II molecules. The designation of their loci consists of three letters: the first (D) indicates the class, the second (M, O, P, Q or R) the family, and the third (A or B) the chain ($\alpha$ or $\beta$, respectively). The individual genes of the HLA system are differentiated by Arabic numbers, and the notation for the numerous allelic variants of these genes is a number preceded by an asterisk. For example, HLA-DRB1*0301 stands for allelic variant 0301 of gene 1, which encodes the $\beta$ chain of a class II molecule belonging to the R family. Class I genes are expressed by most somatic cells, while class II genes are normally expressed by a subgroup of immune cells that include B cells, activated T cells, macrophages, dendritic cells, and thymic epithelial cells. The presentation of short, pathogen-derived peptides to T cells by both class I and class II molecules initiates the adaptive immune response (Klein and Sato 2000a).

Class III (~730 kb of DNA in size), located between class I and class II, encodes at least 62 genes that are likely to be involved in the immune and inflammatory responses (Milner and Campbell 2001). This region is extremely gene dense, with one gene every 10 kb on average and many of the genes are very closely packed. Two clusters in class III
is particularly relevant based on their association with diseases. One is TNF family cluster, a cluster of genes for three related cytokines/cytokine receptors, including tumor necrosis factor alpha (TNFα), lymphotoxin alpha (LTα) and lymphotoxin beta (LTβ). TNFA has been very extensively studied and plays an important role in inflammation, bacterial and viral infection, tumor cachexia and the immune response (Gruen and Weissman 2001). It occurs as membrane bound and secreted homotrimers, and is produced by a variety of cells including predominantly monocytes, macrophages, and some subsets of T cells. LTA (also called TNF beta) has actions that are very similar to TNFα, but less potent than those of TNFα. It is expressed exclusively by lymphocytes. Recent studies suggest a unique role of secreted LTα in the generation of germinal centers, and membrane-bound LTα has been implicated in the development of the spleen, lymph nodes and Peyer’s patches (Milner and Campbell 2001). LTB is a membrane bound molecule that forms a heterotrimer with LTA, therefore activates genes for the immuno-inflammatory response by activating NF-kB. The other is the human MHC complement gene cluster (MCGC), includes complement C2, factor B, RD, SKI2W, DOM3Z, RP, C4, CYP21 and TNX (Yu 1998). This cluster is characterized by polymorphisms, variation in gene size and variable gene number. The genes next to C4, including RP, CYP21 and TNX have been shown to duplicate with C4 and form a genetic module termed RCCX module (Yang and Yu 2000). Three complement proteins C4, C2 and factor B are covered in this cluster, and the role of complement has been discussed in 1.2.
1.3.2 MHC associated diseases

Genetic studies have indicated that genes within the MHC contribute to several hundred diseases, the majority of which are immune-related disorders, e.g. insulin dependent diabetes mellitus (IDDM), rheumatoid arthritis (RA), ankylosing spondylitis (AS), common variable immunodeficiency (CVID) and IgA deficiency (IgAD) (Klein and Sato 2000b, Milner and Campbell 2001). Strong associations have been established between many diseases and alleles of classical HLA class II, and, to a less extent, class I genes. In the case of ankylosing spondylitis, diabetes and rheumatoid arthritis, the reproduction of the disease in transgenic animals expressing particular human HLA antigens strongly indicates that class I or class II molecule itself confers susceptibility to disease (Davidson and Diamond 2001). However, in most cases, it has been not possible to explain the symptoms of the diseases using the variations in the activity of class II or class I polypeptides. Moreover, it is often the extend haplotype, rather than individual alleles, that are associated with disease, suggesting that combinations of alleles across the MHC may predispose to, or provide protection against a disease. Alternatively, some specific MHC class I and class II alleles may associate with diseases due to the strong linkage with the defect in certain class III genes (Milner and Campbell 2001).

The strong linkage disequilibrium of Class I, II and III alleles makes it complicated to definitively analyze genetic factors in the MHC associated diseases. For example, in one of the most common HLA haplotypes, HLA A1 Cw7 B8 DR3 (DRB1*0301 DQB*0201 DPB*0301), which is associated with many autoimmune diseases such as SLE and type 1 diabetes, the C4A gene in class III region is always absent and the -308 position of the TNF-α promoter is always the A-allele. As a result, it
is difficult to distinguish whether it is a specific HLA class II allele, or the absence of
C4A gene, or the presence of -308 A-allele of TNF-α, or an uncharacterized mutation in
one of the 200 identified MHC loci, or a combination of all these factors that contribute
to the disease pathogenesis.

1.3.3 C4 in the MHC-diseases: cause versus linkage disequilibrium

Based on the observation that different polymorphic forms of complement C4
show varying efficiencies of complement activation, it is suggested that susceptibility to
some MHC associated diseases including SLE may be related to the varying efficiency of
complement cell lysis and of immune aggregate dissolution by complement (Porter 1983).
This in turn may depend on the strength of interaction of the different polymorphic forms
of C4 with other proteins, such as C2 and HLA-B in the MHC region, in the scheme of
activation and inactivation of complement. Above average efficiency of complement
activation might increase the risk of tissue damage in autoimmunity and below average
efficiency may slow the dissolution and removal of immune aggregates such as occurs in
SLE. Linkage disequilibrium occurs between alleles for class I, II and III genes and
favorable or unfavorable associations in terms of complement activation may have been a
factor in this linkage.

1.3.4 RCCX modules in human; Similarities and dissimilarities of human and mouse

RCCX in the MHC region

In human MHC class III region, complement C4 is located in a genetic module
termed as RCCX module (Figure 1.3). This module is comprised of four genes, including
a serine/threonine nuclear protein kinase RP, complement protein C4, steroid 21-
hydroxylase CYP21 and extracellular matrix protein tenascin TNX. In an RCCX monomodular haplotype, the intact gene RP1 is always present together with a C4A or C4B gene, a CYP21B gene and a TNXB gene. Duplication or triplication of the RCCX modules occurred by the addition of a genomic segment of either 32.5 or 26.2 kb containing a long or a short functional C4 gene, a pseudogene CYP21A together with the TNXA-RP2 chimera (truncated RP1 and TNXB), i.e. TNXA-RP2-C4-CYP21A (Yu 1998). In human populations, the number of RCCX modules in an individual may vary from 2 to 6. This frequent physical difference in the number and size of the RCCX modules probably contributes to the genetic variability and instability of the MHC class III region. In a patient with juvenile rheumatoid arthritis (JRA-L1), the reciprocal recombination product with two CYP21B genes and a TNXA-XB hybrid in a haplotype was discovered (Rupert et al 1999, Yang et al 1999). The duplicated RCCX modules are highly similar (99% identical) in sequence. There is convincing evidence for the homogenization of polymorphic or mutant sequences in the RCCX modules due to misalignments among monomodular, bimodular and trimodular structures. This homogenization process is not only a driving force for the functional diversities of C4, but it may also be one of the root causes for MHC disease associations (Kawaguchi et al 1991, Yang et al 1999, Yu 1998).

In mouse, there is also a modular gene duplication for RP1/RP2, C4A/C4B (Slp/C4 in mouse), CYP21A/CYP21B, and TNXA/TNXB. However, the RCCX modules have more sequence deviations that include selective integrations of numerous repetitive DNAs, retroelements, multiple point mutations and mini-deletions. These gross sequence changes would decrease the efficiency of alignments between non-allelic genes from
homologous chromosomes. Therefore, homogenizations of the diversified sequences in the duplicated modules would be less frequent in the mouse RCCX than that of human's.

The intergenic sequence between Slp and C4 in the mouse H-2b haplotype is 65.5 kb (Rowen et al 1997b). The orthologous region in humans is 10 kb in size. In humans the RP1/RP2 duplicated sequences are exactly identical, but in the mouse, the duplicated DNA sequence in mouse RP2 contains multiple point mutations and minideletions and is therefore unlikely to be a functional gene. The duplicated region for mouse TNXA is 10 kb with extensive sequence variations/mutations, compared with 4.5 kb and identical sequence for human TNXA. Unexpectedly, a 1.8 kb partially duplicated gene fragment for SKI2W is present between RP2 and TNXA in the mouse, but not in the human MHC. In other words, there is a considerable divergence in the organizations of the human and mouse RCCX (Yu et al 2000).

It is possible that there was a single primordial duplication event involving RP, C4, CYP21 and TNX. Afterwards, the duplicated four-gene modules in human and in mouse took different routes of secondary recombinations, deletions, insertions and point mutations that lead to the current divergence in the RP2 and TNXA gene fragments, and different mechanisms to inactivate one of the duplicated CYP21 genes. It is also of interest to note that in some mouse haplotypes there are multiple copies of Slp/C4 and CYP21 genes (Levi-Strauss et al 1985b, Pattanakitsakul et al 1990, Rosa et al 1985), a phenomenon reminiscent of the C4 polygenic variation in humans.
1.4 Autoimmunity

1.4.1 Central and peripheral tolerance; autoimmune disease

Tolerance is the process that eliminates or neutralizes the autoreactive T and B cells. It is an actively acquired process, rather than an inherited property, in which self-reactive lymphocytes are either prevented from becoming functionally responsive to self antigens or inactivated after encountering these antigens. There are two types of tolerance: central and peripheral tolerance. Central tolerance is generated in primary lymphoid organs including thymus and bone marrow, while peripheral tolerance happens in secondary lymphoid organs such as the lymph nodes and spleen. Autoimmunity results from a failure or breakdown of the mechanisms normally responsible for maintaining self tolerance (Abbas et al 1997).

An autoimmune disease is a clinical syndrome caused by the activation of T cells or B cells, or both, in the absence of an ongoing infection or other discernible cause (Davidson and Diamond 2001). Most autoimmune diseases are complex diseases, caused by the interplay of environmental, hormonal and genetic factors.

The role of environmental factors is predicted by the incomplete concordance of disease expression among monozygotic twin pairs (Salvetti et al 2000). However, in the case of most autoimmune diseases, the actual triggering factor is unkown. One of potential factors is an infectious agent. Microbial antigens can initiate autoreactivity through molecular mimicry, polyclonal activation, or the release of previously sequestered antigens (Davidson and Diamond 2001). In addition, psychological factors such as stress may be a provoking factor for genetically predisposed individuals (Herrmann et al 2000).
The effects of sex hormones in susceptibility to autoimmunity have been studied extensively. Androgen and estrogens can influence the susceptibility and the course of autoimmune diseases. During pregnancy, estrogen and progesterone increase and reach the highest during the third trimester. Intriguingly, in both rheumatoid arthritis and multiple sclerosis, the disease undergoes remission during pregnancy, followed by a flare during the post-partum period. On the contrary, SLE appears to either worsen or remain unchanged during pregnancy (Whitacre 2001). It is indicated that estrogen can alter the B cell repertoire in the absence of inflammation. It appears to suppress cell-mediated and augment humoral-based immunity.

Epidemiological studies have demonstrated that genetic factors are crucial determinants of susceptibility to autoimmune disease. Genetic engineering of mice has led to the identification of at least 25 genes that can contribute to an autoimmune disease when they are deleted or overexpressed (Davidson and Diamond 2001). It is noticed that the defect in a particular gene causes a disease depending on the overall genetic background, and some genetic defects can predispose patients to more than one type of autoimmune diseases, indicating the shared common pathogenic pathways in autoimmune diseases. It is possible that vulnerability of the target organ to immune-mediated damage is also genetically determined. In essence, the predisposition to autoimmune disease represents the net effect of enhancing and protective genes in an interactive network.

1.4.2 C4 and autoimmunity; C4 and self tolerance

Based on the current knowledge, two models have been proposed to explain the role of complement C4 in systemic autoimmunity.
The first is based on the hypothesis regarding the role of C4 in the clearance of immune complexes (ICs) and/or apoptotic debris (Botto 1998, Walport 2000). The classical pathway complement component C1q and C4 can inhibit the formation of large IC lattices and promote the uptake by the mononuclear phagocytic system of the liver through CR1 on erythrocytes (Birmingham and Hebert 2001). Furthermore, Korb and Ahearn observed apoptotic human keratinocytes can directly bind to C1q in the absence of antibodies (Korb and Ahearn 1997), suggesting that C1q may play a role in the recognition of apoptotic bodies. In addition, exposure of phosphatidylserine on the apoptotic cell surface was partially responsible for complement activation and therefore increased the uptake of apoptotic cells in the circulation (Mevorach et al 1998). C4, as downstream of C1q in the classical complement activation pathway, may bind to apoptotic bodies directly or indirectly, therefore helps its clearance from the circulation (Carroll 1998a, Walport 2000). It is demonstrate that mice deficient in C4 spontaneously developed autoantibodies against nuclear antigens, had an initial delay in immune complex clearance (Chen et al 2000) and had defects in the clearance of injected apoptotic thymocytes (Mevorach et al 1998). Guinea pigs deficient in C4 also exhibited delayed clearance of particular immune complexes (Ellman et al 1971).

A second, nonexclusive model is based on the role of C4 in maintenance of self tolerance (Carroll 1998a, Prodeus et al 1998). It is suggested that complement may play a critical role in the transport of self antigens to the bone marrow and may thereby mediate B cell tolerance. Though a study with C4null, Cr2null and C3null mice crossed with mice bearing transgenes specific for both anti-hen egg lysozyme antibody and a soluble form of lysozyme Ag, Michael Carroll’s group demonstrated that C4 and Cr2, but not C3, are
critical in maintaining tolerance to the self Ag (Prodeus et al 1998). In mice, Cr2 codes for both CR1 and CR2 through alternative splicing. The ligands for CR1 include C4 and C3 activation products C4b and C3b, while the ligands for CR2 are C3 split products iC3b, C3dg, and C3d. Further study showed that deficiency for C4 and Cr2, but not C3, in combination with lpr on the C57BL/6 x 129 background resulted in increased severity of lupus-like disease, including renal involvement (Prodeus et al 1998). Autoimmune disease in the C4nullC3null lpr mice is similar in severity to that in C4null lpr mice, but not to that in C3null lpr mice (Einav et al 2002). Moreover, another group found that C4−/− mice, but not Cr2−/− mice, spontaneously produced ANA and DNA specific autoantibodies and exhibited histologic glomerulonephritis in an age- and sex-dependent manner (Chen et al 2000). This may suggest C4 plays multiple roles in systemic autoimmunity and the effect of C4 becomes pronounced with aging and is modified by sex-linked factors. The independence of C4 on Cr2 for systemic autoimmunity suggests that additional receptor for C4 may be involved in this process.

1.4.3 Endogenous retrovirus in autoimmunity

About 5% of the mammalian genome arises through reverse transcription of endogenous retroviruses (ERV). There are many mechanisms by which retroviruses can induce autoimmunity including polyclonal B cell activation, cytokine dysregulation, immune suppression, molecular mimicry, etc (Mountz and Talal 1993). The defect in MRL/lpr mice is caused by the integration of an ETn retrotransposon sequence into the Fas gene leading to abnormal apoptosis and contributing to autoimmunity and lymphocyte accumulation (Mountz and Talal 1993). Genetic studies of spontaneously
autoimmune mice suggest a loss of immune tolerance to ERV *env* protein is responsible for autoimmune glomerulonephritis (Gattoni-Celli et al 1986). Anti-Sm antibodies in lupus patients cross react with the HIV capsid protein p24 (Query and Keene 1987).

1.5 *Systemic lupus erythematosus (SLE)*

SLE is a chronic autoimmune disease, classically depicted as a systemic autoimmune disease caused by the production of pathogenic autoantibodies to a spectrum of nuclear autoantigens. The cause of the disease is multifactorial, including both genetic and environmental factors. It primarily occurs in women (9:1 compared to men) during their childbearing years (Kasharian 1997). The worldwide incidence of SLE is estimated as between 12 and 64 cases per 100,000 individuals, with at least 2- to 4-fold higher incidence in non-Caucasian as compared with Caucasian populations (Kasharian 1997). The manifestation of SLE is diverse with multiple organ systems involved. Approximately one half of lupus patients manifest the severe complications, including nephritis, central nerve system vasculitis, pulmonary hypertension, interstitial lung disease, and stroke (Wakeland et al 2001). The diagnosis of SLE is complicated by the extensive variations in clinical symptoms. It is required that the patient fulfills any four of 11 American College of Rheumatology (ACR) criteria to be diagnosed with the systemic form of SLE (Tan et al 1982). These eleven criteria include: malar or “butterfly” facial rash, discoid rash, photosensitivity, oral ulcers, arthritis with inflammation, serositis (inflammation of tissue around lung and heart), renal disorder (proteinuria or cellular casts), blood disorder (leukopenia, lymphopenia, thrombocytopenia, or hemolytic
anemia), neurologic disorder (seizure or psychosis), immunologic disorder (anti-DNA and anti-Sm antibodies), and abnormal antinuclear antibody titer.

1.5.1 Environmental factors on lupus

Environmental factors have long been considered as triggering factors for lupus development. The suspects include exposure to ultra-violet (UV) light, certain prescription drugs such as heart medicines, antipsychotic drugs, and a few antibiotics for viral and bacterial infection. UV light may cause apoptosis of keratinocytes and promote expression of autoantigens on cell surface membrane blebs, which in turn can lead to high circulating levels of autoantigens and autoreactive B cell expansion (Tsokos and Kammer 2000). The viral infection, such as Epstein-Barr virus (EBV) is a common infectious agent observed in SLE patients. John Harley and colleagues found that an amino acid sequence in EBV that is repeated with one change in an unusual antibody anti-Sm found in about 30% of lupus patients but almost never in people who do not have the disease. However, this result is quite controversial (Marshall 2002).

1.5.2 Sex hormone and lupus

It is well established that sex hormones, such as estrogen, are involved in the pathogenesis of SLE. Estrogens such as estradiol bind to intracellular α and β estrogen receptors (ERα and ERβ), which exert their regulatory effects via binding to distinct estrogen response elements (ERE) in the promoter regions of genes. In T cells of female SLE, estradiol augments expression and activity of calcineurin, a Ser-Thr phosphatase, compared to T cells from normal females, males and lupus males, suggesting estrogen, acting through the estrogen receptors, enhances T cell activation in women with lupus
resulting in amplified T-B cells interactions, B cell activation and autoantibody production (Rider and Abdou 2001). It is also demonstrated that estradiol can break B cell tolerance and induce a SLE-like phenotype in non-autoimmune mice transgenic for the heavy chain of a pathogenic anti-DNA antibody (Tsokos and Kammer 2000).

1.5.3 Genetic defect in human SLE

SLE is a complex genetic disease, which involves many genes contributing to different aspects of disease manifestation. Linkage studies using sib pairs and extended pedigrees have suggested seven loci in the human genome are significantly linked to SLE (LOD score >3.3), including 1q23, 1q25-31, 1q41-42, 2q35-37, 4p16-15.2, 6p11-21, and 16q12 (Tsao 2003). Detailed sequence analysis and mouse models suggest several candidate genes fit within these loci. Specifically, Fcγ receptor genes FCGR2A, FCGR2B and FCGR3A are located in 1q23. MHC Class I, class II and complement C4 and C2 are located in 6p11-21. Poly (ADP-ribose) polymerase (PARP) is at 1q41-42, and PDCD1 at 2q35-37. The current understanding of the role of these genes and some other genes with related functions are discussed as follows.

1.5.3.1 Complement genes and genes involved in disposal of cellular debris and immune complexes

Complement genes, including C1q, C4 and C2 are among the first set of identified genes to be related with SLE. Complete deficiencies of C1q, C4 and C2 are extremely rare, but the prevalence of lupus in these individuals are >90%, >90% and 33%, respectively (Atkinson and Schifferli 1999). In mixed B6/129 genetic background, mice with targeted deletion of C1qa showed glomerulonephritis with immune deposits and
apoptotic cells in the glomeruli at one year of age (Botto et al 1998). With the same genetic background, all C4-deficient females and most males produced antinuclear antibodies or lupus-like phenotypes by 10 months of age (Chen et al 2000). In addition, it is also suggested that defects in complement receptors 1 and 2 and mannose binding protein are involved in lupus development. The mechanism of the involvement of complement proteins in SLE underlies the role of complement in disposal of cellular debris and immune complex. Besides complement proteins, two other susceptibility genes, Dnase 1 and SAP (serum amyloid P component) have been identified based on their role in the removal of autoantigens. Polymorphisms in the Dnase 1 gene have been detected in SLE patients, suggesting that polymorphism may potentiate SLE in humans (Yasutomo et al 2001).

1.5.3.2 HLA

The HLA region at chromosome 6p21.31 contains 224 identified loci within 3.6 Mb. The HLA-DR2 and DR3 class II genes have consistently associated with SLE in many studied Caucasian populations with a 2- to 3-fold increase in the frequency of these two alleles compared with controls (Tsao 2003). Given the role of class II molecules in the T-B cell interaction, it is not surprising that certain DR/DQ alleles are strongly associated with particular IgG autoantibody profiles in SLE. The strong linkage disequilibrium within the MHC region results in extended haplotypes, which makes it difficult to interpret these HLA associations and ultimately localize the genetic defect (Tsao 2003, Wakeland et al 2001). Using genotypes of the MHC microsatellite markers from ~300 families, three class II containing SLE risk haplotypes (DRB1*1501(DR2)/DQB1*0602, DRB1*0801(DR8)/DQB1*0402 and
DRB1*0301(DR3)/DQB1*0201) have been identified by the Transmission Disequilibrium Test (TDT) analysis. The estimated genotype relative risk for each of the three haplotypes is between 1.9- and 2.6-fold and individuals with two copies of risk haplotypes exhibit a dose-dependent increased risk (3.5 to 6.8 fold) for SLE (Tsao 2003).

1.5.3.3 Fcγ receptor genes FCGR2A, FCGR3A, FCGR3B and FCGR2B

The genes that encode for FcγRII (CD32) and FcγRIII (CD16) function to bind and clear IgG antibodies and IgG-containing immune complexes from the circulation. They are members of the immunoglobulin gene superfamily, clustered within 100 kb at 1q23. Because SLE is a disease characterized by tissue deposition of IgG containing immune complexes, it is possible that genetic polymorphisms of these Fc receptor genes would be associated with lupus. It is shown that a greater portion of African American SLE patients are heterozygous or homozygous for the IgG2 low binding FcγRIIa-R131 allele, with further enrichment observed in patients with lupus nephritis (Salmon et al 1996). However, the results from other ethnic groups are quite controversial (Duits et al 1995, Karassa et al 2002, Smyth et al 1997). A low binding allele of FcγRIIIa, which functions to bind and clear IgG1- or IgG3-bearing immune complexes, is also associated with SLE nephritis in patients of European, African and Asian, but not for SLE susceptibility per se in the absence of nephritis (Karassa et al 2003). Because IgG2 and IgG3 are major subclasses of ICs deposited in renal biopsies of lupus nephritis patients, it is suggested that the relative importance of FcγRIIa and FcγRIIIa polymorphisms on lupus may depend on the IgG subclass of pathogenic autoantibodies (Tsao 2003).
1.5.3.4 Apoptosis factors

Apoptosis, the process of programmed cell death, has been linked to lupus pathogenesis. As antigen, high circulating levels of apoptotic materials may stimulate autoreactive B cells to produce autoantibodies, leading to immune complex formation and tissue deposition. On the other hand, as immune modulator, impaired apoptosis in T cells may allow autoreactive cells to mature that predisposes patients to diseases. The relationship of apoptosis to lupus, however, can’t be fit into a simple patitude. Roles of both “too much” and “too little” apoptosis have been found (Greidinger 2001).

Of the candidate genes in the 1q41-42 region, poly(ADP-ribose) polymerase (PARP), a Zinc-finger DNA-binding protein involved in DNA repair and apoptosis, were preferentially associated with affected offsprings in a large number of families. Autoantibodies that bind to the Zinc-finger motifs of PARP are frequently found in patients with autoimmune diseases, which may prevent caspase-3-mediated PARP cleavage during apoptosis. However, the association of PARP with lupus could not be found in French and German Caucasian studies (Tsao 2003, Tsokos and Kammer 2000).

Fas, a receptor of the TNF-receptor family, is known to induce apoptosis when activated by antibodies against it. Fas ligand is a membrane bound member of the TNF family. It could also induce cell apoptosis when bound to Fas on target cell membrane. MRL mice with defective Fas genes (MRL/lpr) or with defective Fas ligands (MRL/gld) develop a lupus-like syndrome with extremely high penetrance (Greidinger 2001). It is noticed that freshly isolated B cells and CD4+ and CD8+ T cells from lupus patients have higher levels of Fas antigen and higher rates of spontaneous apoptosis compared to
normal cells. Defects in the genes encoding the Fas and Fas ligand have been identified in individual SLE patients (Tsokos and Kammer 2000).

Other defects in immune system apoptosis linked to lupus include a recently described circulating factor, BlyS (BAFF, TALL-1, THANK, zTNF-4). It is shown to dramatically increase susceptibility in animals when overexpressed. BlyS is also overexpressed in human lupus compared to controls. It is found that increased BlyS in SLE patients was partially associated with higher levels of anti-dsDNA antibody of the IgG, IgM, and IgA classes, but not associated with the disease activity. A function of BlyS is as a B cell-specific anti-apoptotic mediator, therefore, the association of BlyS overexpression and lupus could be due to impaired B cell apoptosis (Greidinger 2001, Zhang et al 2001).

1.5.3.5 Lymphokines

SLE is thought to be Th2-mediated autoimmune disease. Previous studies suggest that cytokines TNFα, IL-10 and IL-6 may be related to lupus development.

In lupus patients, a shift from a balance between Th1 and Th2 to Th2 dominance results in the overexpression of IL-6 and IL-10 cytokines and underproduction of IL-2 and IFN-γ. In T cells obtained from lupus patients, the mRNA levels for IL-6 and IL-10 are increased (Tsokos and Kammer 2000).

TNFA gene (gene for TNFα) is located in MHC class III region, close to C4 and C2. It is a cytokine that plays a role in the pathogenesis of many autoimmune diseases. The TNFA gene exhibits a promoter polymorphism at –308. The -308A allele is linked predominantly to HLA-DR3 and C4A deficiency. In studies of Taiwanese (Lu et al 1997)
and African Americans (Sullivan et al 1997), the -308A allele is associated with increased risk for SLE. However, in South African African Americans, there is a reduction of this allele in the patients (Rudwaleit et al 1996).

1.5.3.6 PDCD1 (PD-1) gene

PDCD1, located at 2q35-37, is an inhibitory immunoreceptor that has a pivotal role in peripheral tolerance. Mice deficient in PDCD1 develop glomerulonephritis and arthritis in C57BL/6 background. Intriguingly, a promoter polymorphism in PDCD1 is associated with lupus in a recent study of ~2500 individuals (Tsao 2003).

1.5.4 Genetic insight into murine lupus

Murine systemic autoimmunity has been recognized as a model for human lupus for over 30 years. The disease is extensively studied in several inbred strains and synthetic models created by targeted gene disruptions and transgenes. It is suggested as the most useful models for the analysis of genetic mechanisms, especially those caused by single genes. Three spontaneous lupus-prone models have been genetically characterized in details. They are 1) the NZBxNZW hybrid and the related NZM2410 congenic recombinant strain, 2) the MRL/lpr strain, and 3) the BxSB strain. NZWxNZB is considered as the best model for human SLE, in that it shows a strong female predominance in susceptibility, develops a severe lupus with immune complex mediated fatal glomerulonephritis, and displays a wide variety of anti-DNA antibodies (Rudofsky and Lawrence 1999). The MRL/lpr strain that carries the lpr mutation of Fas gene on an autoimmune-prone MRL background, also develops lupus with nephritis. The Fas mutation is a key element in the penetrance of the disease in this murine model; however,
it is not associated with lupus development in human. There is a slight preponderance for female in this model (Nose et al 2002). The susceptibility to lupus in BxSB strain is limited to male because of the presence of a Y chromosome carrying yaa that interacts with several genes in the BxSB genome (Izui et al 2000). The male BxSB develops severe lupus, and then dies of progressive nephritis. The analysis of these models indicates that murine lupus is also genetically complex and mediated by combinations of genes. Although each lupus-prone strain is susceptible in part due to a unique set of disease genes, several of the susceptibility loci mapped to similar locations across multiple strains have been identified, including specific regions of chromosome 1, 4, 7 and MHC on chromosome 17. Further studies using congenic dissection helps identifying the loci and some specific genes linked to special phenotypes observed in lupus. Moreover, the function of individual genes in lupus development can be evaluated by gene knockout and transgenic approaches. So far, four broadly defined pathways impacting immune functions relevant to autoimmunity and the susceptibility genes in each group have been identified. They include: 1) genes related to immune complex and apoptotic body clearance, e.g. C1q, C4, C3, SAP, Dnase, Serum IgM, Fcγ and Mer; 2) genes involved in lymphoid signaling, e.g. SHP-1, Lyn, CD22, BlyS, PD-1, IL-2, CD45, G2A, and IFNγ; 3) genes related to apoptosis, e.g. Fas, Fas-L, Bcl-2, Pten, and P21; 4) genes in epitope modification such as α-mannosidase II. However, some genes highly related to murine lupus have not been confirmed in humans (Wakeland et al 2001).

Although the murine model is useful and powerful in lupus study, numerous differences exist between mice and humans that suggest that mouse studies are not always applicable to human disease. Several differences in disease manifestation between
mice and humans are noticed. First, human lupus onset is gradual, slowly progressive and many patients only have mild disease without complications; while mouse lupus is acute and always lead to end organ injury such as nephritis and renal failure. Second, the influence of environmental factor on disease is much stronger in human lupus than that in mouse lupus. Third, the genetic penetrance of lupus is very strong in mice, that is, almost every mouse with the suggested defect develops lupus; while it is relatively weak in humans. In addition, the differences in the biological systems that interact with circulating immune complex, and in the general physiology between two species also suggest additional genes unique or functionally different in humans may contribute to lupus development in humans (Birmingham et al 2001).

1.5.5 A three-step hypothetical model of lupus

Based on congenic dissection of NZM2410 lupus-prone mouse strain into a B6 background and gene knockout and transgenic mice studies, a three-step hypothetical model of the roles that various genetic pathways play in the initiation of SLE pathogenesis is presented (Wakeland et al 2001). The development of SLE can be viewed as involving interactions between genes in these three pathways. The first pathway contain genes such as sle1, C1q, C4, SAP, and DNase 1 that can trigger the loss of immune tolerance to nuclear autoantigens and mediate the initiation of autoimmunity. At the second step, genes as sle2, sle3, Fas, FasL, yaa, lyn, BlyS, IL-2 and PD-1 participate in the process of hypersensitization and/or disruption of the normal immune system, leading to immune dysregulation. Finally, FcγIIIa, sle6, and sle1d are genes that would be involved in the final step mediating ANA targeting into specific organs to promote end organ damage.
1.6 C4 and human SLE

It was observed about fifty years ago that low serum complement activity or low protein concentrations of complement C4 concurred with disease activities of systemic lupus erythematosus (SLE). Complete deficiencies of complement components C4A and C4B, albeit rare in human populations, are among the strongest genetic risk factors for SLE or lupus-like disease, across HLA haplotypes and racial backgrounds. However, whether heterozygous or partial deficiency of C4A (C4AQ0) or C4B (C4BQ0) is a predisposing factor for SLE has been a highly controversial topic.

1.6.1 Changes of complement activity and C4 protein concentrations in SLE

It has been long known that SLE patients manifest reduced complement hemolytic activity (CH50) (Elliot and Mathieson 1953, Vaughan et al 1951). In addition, reduced serum or plasma levels of complement C1q, C4 and C3 have been consistently observed in lupus patients, particularly those with lupus nephritis (Cameron et al 2003, Lehman et al 1979, Lewis et al 1971). Serial analysis of serologic factors in SLE revealed that in many patients lower C4 levels occurred before the depression of other complement components. After the induction of remission, C4 had a tendency to return to normal levels more slowly than C3 (Cameron et al 2003, Gewurz et al 1968, Kohler and Ten Bensel 2003, Lloyd and Schur 1981, Swaak et al 1986). Depressed levels of serum or plasma C3 were found in a variety of glomerulonephritis patients but depressed C4 levels were mainly present in lupus nephritis. The lower C4 and C3 levels in active disease imply the presence of anticomplementary activities, a decreased biosynthesis, or an increased catabolism through complement activation by immune complexes. Why the
low C4 levels persist in many lupus nephritis patients after disease remission is not understood.

Serial monitoring of plasma or serum C4 and C3 concentrations and CH50 has been used as a guide to the clinical management of lupus or lupus-like diseases (Appel et al 1978, Cameron et al 1973, Hebert et al 2001, Ricker et al 1991, Schur 1978). In a four-year serial study of 10 SLE patients in Texas, Moulds and colleagues showed that parallel and proportional changes of serum C4A and C4B protein levels occurred with changes in SLE disease activities, suggesting that there is no differential metabolism between serum C4A and C4B proteins in the SLE patients studied (Moulds et al 1993). A prospective, monthly study of 21 Swedish SLE patients without nephritis (Mollnes et al 1999) concluded that results of the complement tests were unable to predict disease flares of SLE without nephritis (Mollnes et al 1999). It was observed in this study that slight to moderate decreases in plasma C4 levels was found in 10 patients (47.6%) and severely decreased C4 levels were present in 7 patients (33%). However, C4 concentrations were low during the entire observation period in half of the patients and the low levels were not associated with increased complement activation, suggesting a genetic defect in C4 synthesis. In a longitudinal, monthly study of 53 general lupus patients in Baltimore, depressed levels of C4 and C3 were observed with a concurrent increase in renal and hematologic disease activities in SLE. However, decreases in C4 and C3 levels were not consistently associated with SLE flares (Ho et al 2001).

As laboratory tests of serum or plasma C4 and C3 concentrations and CH50 are unable to accurately predict SLE disease flares, innovative approaches are explored by assaying complement activation or inactivation products in plasma and in urine.
Complement split products represent complement activation and inactivation triggered by the presence of immune complexes or apoptotic materials, therefore it may better reflect the disease processes. An early study using Futhan-EDTA plasma for radioimmunoassays (RIA) of complement anaphylatoxins from 24 SLE patients revealed that 20 patients (83%) had elevated levels of C4a, with significantly higher levels found in 16 patients who had severe disease and required azathioprine treatment in addition to prednisolone. Unexpectedly, a parallel increase in C3a levels was not observed (Wild et al 1990). In a prospective 15-month study with 380 plasma samples from 86 SLE patients, Buyon and colleagues (Buyon et al 1992) found that C4d was most sensitive with regard to subsequent disease flare, while Bb was most specific and predictive. In end-stage renal disease patients, an elevated level of plasma C4d and a depressed level of native C4 were only observed in lupus nephritis patients but not in patients without SLE (Buyon et al 1992). Subsequently, Manzi and colleagues (Manzi et al 1996) found that the complement split products C4d and Bb in plasma were sensitive indicators of moderate to severe lupus disease activity, and were more informative in situations when C4 and C3 levels remained normal despite evidence of clinical disease activity. Urine C3d were found in three types of acute nephritis patients engaged in the study, although urine C3d was present in some SLE patients without evidence of proteinuria.

C4d, present in soluble and membrane-bound forms, is a proteolytic product of the C4b generated by factor I cleavage in the presence of a cofactor (C4b binding protein). The membrane-bond C4d fragment is not assayable in plasma or urine samples. Therefore, assay of the soluble C4a has its merits by better representing the entire process of C4 activation. In summary, measurement of the plasma levels of complement C4, C3
and their split products provides relevant immunologic parameters that help assessing and predicting disease activities of SLE. A long serial study period from a large cohort of renal and extrarenal SLE patients with a well-defined genetic background including complement C4A and C4B gene dosages would be highly desirable.

1.6.2 Complete C4A and C4B deficiencies in SLE and immune complex diseases

To date 26 humans from 18 families with complete complement C4 deficiency (C4D) have been firmly established. Among them, 14 individuals (9 females and 5 males) were diagnosed with SLE according to the ACR criteria (Tan et al 1982), seven had a lupus-like disorder such as photosensitive skin lesions and/or discoid lupus, and four had kidney diseases such as mesangiproliferative glomerulonephritis, recurrent hematuria, membranous nephropathy, and Henoch-Schonlein purpura with end-stage kidney failure (Lhotta et al 1993b, Lhotta et al 1993a). Only one subject, age 21 at the time of report, remained relatively healthy.

The age of SLE disease onset/diagnosis among the C4 deficient subjects varied from 2 to 41 years old. Four C4 deficient patients died at ages between 2-25 years old. Common classical manifestations of C4 deficiency are photosensitivity, severe skin lesions, digital erythema, Raynaud’s phenomenon, anti-Ro/SSA, rheumatoid factor and recurrent bacterial and viral infections.

1.6.2.1 Immunogenetics of C4 deficiency

At least 15 different HLA haplotypes have been found in C4 deficient patients, of whom 19 (73%) were homozygous with the same HLA haplotypes on both chromosomes. Four HLA haplotypes with complete C4 deficiency were detected in
multiple individuals and families. HLA A2 Cw3 B40 (DR6), the most common C4 deficiency haplotype, has been shown in six different Caucasian patients from Europe and North America. TaqI southern blot analysis of SLE patients with this haplotype revealed a monomodular long RCCX structure consisting of a mutant C4A gene (Fredrikson et al 1998, Lokki et al 1998, Uring-Lambert et al 1989, Welch et al 1990). The molecular defect leading to the non-expression of C4A protein has been elucidated from two Scandinavian families (Fredrikson et al 1998, Lokki et al 1998). A 2-bp (TC) insertion at the sequence for codon 1213 in exon 29 is present in the mutant C4A genes, leading to frame-shift and non-sense mutations. This HLA haplotype with complete C4 deficiency appears to be related to HLA haplotype A2 Cw3 B60 DR6 with a bimodular LS structure where the same mutation happens in the long C4AQ0 gene (Braun et al 1990, Madi et al 1991). The short C4B gene from the HLA B40 haplotype with RCCX-LS could have been lost through an unequal crossover. Alternatively the gene conversion-like events could have spread the 2-bp insertion to the other C4 genes. Indeed, it has been found that the mutant C4A gene from the HLA A2 B12 DR6 (Rupert et al 2002), and the mutant C4AQ0 and C4BQ0 genes in HLA A2 Cw7 B39 DR15 (Lokki et al 1998) also possess the identical molecular defect. Besides the 2-bp insertion at codon 1213 from exon 29, two other deleterious mutations have been found in mutant C4 genes, which are single C-nucleotide deletions at codon 522 from exon 13 of the C4BQ0 gene for the HLA A2 B12 DR6 haplotype, and at codon 811 from exon 20 of the C4AQ0 gene in the HLA A30 B18 DR3 haplotype.

In the European Alpine region, homozygous haplotypes with HLA A30 B18 DR7 were present in six subjects from three different families, and HLA A24 Cw7 B38 DR13
was present in four C4-deficient individuals from two related families. The molecular
defects of the mutant C4 genes from these two haplotypes are being characterized.

To date all mutant C4 genes characterized were from Caucasian families. Unfortunately, mutant C4 genes from other ethnic groups such as Orientals or Africans have not been available for characterization.

1.6.2.2 Impairments of immune response in C4 deficient patients

An early study of the immunologic parameters in a Caucasian SLE patient with
heterozygous HLA A2 B12 Dw2 / A2 Bw15 Dw8 and complete C4 deficiency revealed a
depressed primary immune response when immunized with bacteriophage φX174,
abnormal immunologic memory, and failure to switch from IgM to IgG during secondary
response. Persistent lymphopenia and reduction in lymphocyte response to mitogen
stimulation was also documented (Jackson et al 1979). In addition, the opsonization
activity of zymosan and the neutrophil chemotactic activity by the C4 deficient serum
were markedly impaired, when compared with those from healthy controls (Clark and
Klebanoff 1978). In another study of two C4 deficient Moroccan siblings with lupus-like
disease and persistent pulmonary infections, diminished opsonization and bactericidal
activities by the C4 deficient sera were observed. However, in other aspects the humoral
and cellular immunity appeared normal (Mascart-Lemone et al 1983). These two girls
had homozygous HLA haplotype A11 B35 Cw4 DR1. Such difference in the phenotypic
manifestations underscores the relevance of genetic backgrounds and racial difference in
immunity (Burchard et al 2003).

In essence, complete C4A and C4B deficiencies appear to be a strong genetic risk
factor for lupus, lupus-like, or immune complex-related kidney diseases. The disease
severities and the organ involvement varied among the C4 deficient patients, perhaps because the genetic backgrounds including the HLA haplotypes were quite diverse. Some C4 deficient subjects may have protective factors that reduce the negative effects of C4 deficiency, as exemplified by one complete C4 deficient subject who was healthy at least to age 21 years (Lhotta et al 1993b).

1.6.3 Epidemiological study of C4 and human SLE

In the early eighties, two research groups, one led by Richard Bachelor in London, UK and the other by Roger Dawkins in Perth, Australia, reported the association of C4 null alleles (C4Q0) or partial deficiency of C4A and/or C4B with human SLE (Christiansen et al 1983, Fielder et al 1983). Since then the associations of C4AQ0 or C4BQ0 with SLE have been examined extensively in many ethnic groups. Most studies revealed significant associations of homozygous and heterozygous C4AQ0 in the SLE patient groups, when compared with race-matched controls of healthy individuals from the same geographical locations. However, some investigators vigorously challenge the reported association of C4AQ0 or C4BQ0 with SLE (Dragon-Durey et al 2001, Fronek et al 1990, Hartung et al 1992, Schur et al 1990). The most common argument has been that C4AQ0 was simply a marker for linkage disequilibrium with another gene in the MHC that was responsible for SLE, which could be located close to HLA DR3 or DR2. Another argument was that the frequency of C4AQ0 in some SLE population was close to that of pooled controls with about two thousand haplotypes (Fronek et al 1988, Schur et al 1990).

Cumulative results from more than thirty-five different studies revealed that heterozygous and homozygous deficiencies of C4A (C4AQ0) were present in 40-60% of

French SLE and control populations had relatively low frequencies of C4AQ0 but the difference between the patient and control groups was statistically significant (Cornillet et al 1993, Goldstein and Sengar 1993). The C4AQ0 allelic frequency is only 3.7% in the French control population, and 16.9% in the French SLE patients. The latter is very close to the C4AQ0 frequency in the healthy controls of Anglo-Saxons (16.9%) and Nordic (14.1%). These results emphasize the importance of applying the appropriate ethnic group and matched controls for statistical analyses. The relative risk of C4AQ0 in SLE varied between 2.3-5.3 among different ethnic groups.

In Caucasian and African SLE patients, the two major causes for C4AQ0 are (1) the presence of a mono-S RCCX module with a single, short C4B gene in the major histocompatibility complex (MHC); and (2) a 2-bp insertion into the sequence for codon 1213 at exon 29 of the mutant C4A gene. However, at least nine different HLA haplotypes, many without mono-S linked to HLA DR3, have been demonstrated in
Icelandic SLE patients with the C4AQ0 (Kristjansdottir et al 2000, Steinsson et al 1995, Traustadottir et al 1998). The mono-S in Africans and African Americans is frequently associated with haplotypes such as HLA B44 DRB1*1503 (DR2) instead of B8 DR3 (Fraser et al 2000). Furthermore, both mono-S structures and 2-bp insertion in exon 29 are absent in the C4AQ0 of Oriental SLE patients. The highly significant association of C4AQ0 with SLE across multiple HLA haplotypes and ethnic groups, and the presence of different mechanisms leading to a C4A protein deficiency among SLE patients suggested that deficiency or low expression level of C4A protein is a primary risk factor for SLE disease susceptibility *per se*. However, such a conclusion does not exclude the possibility that other genetic element(s) present in the MHC could also contribute to the risk of SLE. Further analysis of British SLE patients revealed that there may be two primary risk factors from the MHC, one is C4AQ0 (Green et al 1986), and the other is associated with HLA DQB1*0501, which is in strong linkage disequilibrium with DRB1*0301 (Davies et al 1995). Besides the 8.1 AH, there is also a high prevalence of HLA B7 DR2 (DRB1*1501) C4A3 C4B1 (LL) in the German SLE (Hartung et al 1992). Indeed the relevance of DR2 in SLE has been inferred by multiple studies (Doherty et al 1992, Hawkins et al 1987, Hong et al 1994, Howard et al 1986, Olsen et al 1989, Reveille et al 1995). In a study of black SLE, Howard and colleagues observed C4AQ0 and DR2 each contributed a relative risk about 3 but their presence together *in trans* increased the relative risk to 24.9. Reveille and colleagues showed that the DR2 in African American SLE patients belongs to DRB1*1503 (Reveille et al 1991).

Intriguingly, Spanish, Mexican, Australian Aborigine SLE patients had increased frequencies of C4B deficiency instead of C4A deficiency (Christiansen et al 1991, Juan
Such observations underscore the importance of both C4A and C4B proteins in the fine control of autoimmunity. Different racial and genetic backgrounds could change the thresholds for the requirement of C4A or C4B protein levels in immune tolerance and immune regulation.

Despite the high prevalence of heterozygous C4A deficiency in SLE patients in almost all ethnic groups except Spanish and Mexicans who have high frequency of C4B deficiency, a corresponding increase in homozygous deficiency of C4A is only observed in specific populations. In Caucasians of northern European origin, a frequency of 5-14% in homozygous C4A deficiency is observed in SLE patients, compared with 1-2% in the control populations. A similar frequency has been found in Hans Chinese residing in mid-China (Zhao et al 1989) and southwestern China (Man et al 2003) and in Japanese. However, many other studies detected a very low frequency of homozygous C4A deficiency in SLE patients in Orientals including Southern Chinese residing in Hong Kong (Dunckley et al 1987) and in Kuala Lumpur of Malaysia (Doherty et al 1992), Koreans (Hong et al 1994) and Japanese (Yamada et al 1990). Caucasian and African American SLE patients with homozygous C4A deficiency tend to have milder disease that includes a lower incidence of renal and CNS disease. By contrast, Chinese SLE patients with homozygous C4AQ0 appear to have more severe disease as shown by high incidence of renal disease, serositis and high levels of anti-dsDNA (Man et al 2003).
1.7 Goals of study

Complement C4, located in MHC class III region, functions in direct lysis of invaded microbes by activating the complement pathways. C4 can also help the clearance of microbes by phagocytes and the removal of immune complexes through a mechanism called as opsonization. It is suggested that C4 deficiency may lead to defective processing of immune complexes, impairment of B cell memory, and persistence of bacterial and viral infections. On the contrary, over-production or over-activation of C4 may exacerbate the local inflammation and tissue injuries.

Human C4 is one of the most polymorphic proteins in the serum with a wide range of protein levels among different individuals and ethnic groups (100 -1000 µg/ml). The great genetic diversity of human C4A and C4B genes could lead to quantitative and qualitative variations of the polymorphic C4A and C4B proteins, which may be one of the intrinsic mechanisms of response to selection against a great variety of microbes or parasites. However, the deeply ingrained “two-locus” model of human C4 genes greatly hindered its epidemiological study in humans. The purpose of this dissertation was to accurately determine C4 polygenic and gene size variations, C4 mutations, protein polymorphism and protein levels in humans including normal ethnic groups and C4-related disease patients, particularly, SLE patients. More specifically, three objectives were proposed:
1) To determine the genetic diversity and protein polymorphisms of human complement components C4A and C4B in healthy Caucasians from a specific Caucasian population in Central Europe

It was first observed in Ohio Caucasians that there were different frequencies of C4A and C4B allotypes and/or different RCCX length variants with one, two or three long or short C4 genes. However, the variations of gene dosages and gene size in other populations and their effect on protein levels and functions of C4A and C4B are unknown. Therefore, the first objective of this dissertation was to determine C4 genotypic and phenotypic variations, both qualitatively and quantitatively, in healthy subjects from a different Caucasian population. The underlying hypothesis is that the variations of gene dosages and gene size in C4 will affect its protein levels and protein function, therefore affect the intrinsic strength of the human complement system. As C4 is an important molecule in innate immunity, determination of genetic diversity of C4 in another Caucasian population will help validate our earlier observation of its polygenic variation in Ohio Caucasians (Blanchong et al 2000). In addition, elucidation of the effect of gene number and size on the protein expressions of C4A and C4B may provide a foundation for the quantitative traits in innate immunity.

2) To elucidate the molecular basis of complete complement C4 deficiency in human SLE and kidney disease patients

Deficiencies of complement C4A and/or C4B are associated with a variety of autoimmune and infectious diseases. All but one HLA-typed, C4-deficient individuals experienced symptoms related to immune complex or clearance disorders such as lupus or lupus-like disease, nephritis or kidney disease. In the 26 documented complete C4
deficiency subjects, the molecular basis of C4 deficiency was determined only in five subjects. The underlying hypothesis was that the C4 deficiency was caused by gene deletion or nucleotide mutation(s) in the C4A and C4B genes. Determination of the nucleotide mutations in C4 gene will help epidemiological study of the association of the polygenic variations of C4 with a variety of infectious and autoimmune diseases.

3) To determine complement C4A and C4B diversities, RCCX modular variations and risk factors of MHC class III genes in Ohio SLE

SLE is a prototype of systemic autoimmune disease. The role of C4 in SLE has long been proposed. However, most previous studies used the phenotypic analysis dependent on the two-locus model, an accurate account of polygenic variation of C4A and C4B has never been performed. In addition, C4 is in the highly gene dense MHC region, where specific alleles from different genes are linked together, therefore it is difficult to evaluate the contribution of individual genes to the disease manifestation. Determination of C4A and C4B gene dosage, C4 protein polymorphism and protein level in the SLE patients will allow us to examine the effect of C4 deficiencies, overdosage and/or over-expression on SLE more precisely. The underlying hypothesis is that deficiencies of C4 or C4A lead to inefficiency in clearance and deposition of IC in tissues, while high C4 protein levels in the circulation and/or in local tissues, caused by high C4 gene dosage or high C4 protein expression, over-activate the complement pathways and aggravate tissue injury. To evaluate the contribution of other factors, the variations in two other candidate genes C2 and TNFA in MHC class III region will be studied. Elucidation of gene dosage variation of total C4, C4A and C4B in Caucasian and in African
American patients with SLE, will yield valuable information regarding to diagnosis and therapeutic intervention of this chronic disease.
Figure 1.1 The Three Activation Pathways of Complement: the Classical, Mannose-Binding Lectin, and Alternative Pathways. The three pathways converge at the point of cleavage of C3. The classical pathway is initiated by the binding of the C1 complex to antibodies bound to an antigen on the surface of a bacterial cell. The mannose-binding lectin pathway is initiated by binding of the complex of mannose-binding lectin and the serine proteases mannose-binding lectin–associated proteases 1 and 2 (MASP1 and MASP2, respectively) to arrays of mannose groups on the surface of a bacterial cell. The alternative pathway is initiated by the covalent binding of a small amount of C3b to hydroxyl groups on cell-surface carbohydrates and proteins and is activated by low-grade cleavage of C3 in plasma.
Figure 1.2 A molecular map of the genes in the MHC complement gene cluster. Horizontal arrows represent gene configurations.

Figure 1.3 The polygenic and gene size variations of human complement C4 and RP-C4-CYP21-TNX (RCCX modules) in the MHC class III region.
Figure 1.4 The qualitative variations of complement C4A and C4B proteins in the plasma based on gross charge difference resolved by high voltage gel electrophoresis. The basic C4B allotypes usually move slower than the acidic C4A allotypes. The red bars indicate the most common C4A and C4B allotypes.
Figure 1.5 Protein structure of human C4. A. Structural features and polymorphisms of human C4A and C4B proteins. B. C4d crystal structure of the C4A isotype resolved at 2.3 angstrom. C. Molecular surface representation of the concave surface of C4Ad.
Figure 1.6 The sophisticated genetic diversity of C4 genes. **a.** The polygenic variation of human C4 genes. The Pmel digested fragments of mono-, bi-, and tri-modular RCCX structures with one, two and three copies of C4 genes, respectively, are indicated. **b.** The dichotomous size variation of human C4 genes. **c.** Restriction enzymes recognizing the DNA sequences specific for C4A and C4B isotypic residues. **d.** TaqI RFLP to determine the number and size of C4 genes and RCCX modules. **e.** Pulsed field gel electrophoresis (PFGE) of Pmel-deigested genomic DNA to elucidate the number of RCCX modules.
A. The 1-2-3 loci concept of human C4 genes in the MHC

<table>
<thead>
<tr>
<th>LLL</th>
<th>RP1</th>
<th>C4A</th>
<th>C4B</th>
<th>TNXB</th>
</tr>
</thead>
<tbody>
<tr>
<td>LLL</td>
<td>RP1</td>
<td>C4A</td>
<td>C4B</td>
<td>TNXB</td>
</tr>
<tr>
<td>LL</td>
<td>RP1</td>
<td>C4A</td>
<td>C4B</td>
<td>TNXB</td>
</tr>
<tr>
<td>L</td>
<td>RP1</td>
<td>C4A</td>
<td>C4B</td>
<td>TNXB</td>
</tr>
</tbody>
</table>

Pme I fragment (kb): 178, 146, 113, 107

long RCCX module: 32.7 kb
short RCCX module: 26.3 kb

B. The dichotomous size variation of human C4A and C4B genes

Long C4 gene, 20.6 kb

Short C4 gene, 14.2 kb

Intron 9

C. Restriction enzymes recognizing the nucleotide changes specific for C4A and C4B

D. Taq I RFLP

E. Pme I PFGE

Figure 1.6 (From Yang, Y. et al, 2004 Curr Dir Autoimmun 7:98-132)
Figure 1.7 A comparison of frequencies of homozygous (A) and heterozygous (B) C4AQ0 in SLE patients and matched controls in various ethnic groups/races.

(From Yang, Y. et al, 2004 Curr Dir Autoimmun 7:98-132)
2.1 ABSTRACT

Among the genes and proteins of the human immune system, complement component C4 is extraordinary in its frequent germline variation in the size and number of genes. Definitive genotypic and phenotypic analyses were performed on a central European population to determine the C4 polygenic and gene size variations and their relationships with serum C4A and C4B protein concentrations and hemolytic activities. In a study population of 128 healthy subjects, the number of C4 genes present in a diploid genome varied between two to five, and 77.4% of the C4 genes belonged to the long form that contains the endogenous retrovirus HERV-K(C4). Intriguingly, higher C4 serum protein levels and higher C4 hemolytic activities were often detected in subjects with short C4 genes than those with long genes only, suggesting a negative epistatic effect of HERV-K(C4) on the expression of C4 proteins. Also, the body mass index appeared to affect the C4 serum levels, particularly in the individuals with medium or high C4 gene dosages, a phenomenon that was dissimilar in several aspects from the established
correlation between body mass index and serum C3. As expected, there were strong, positive correlations between total C4 gene dosage and serum C4 protein concentrations, and between serum C4 protein concentrations and C4 hemolytic activities. There were also good correlations between the number of long genes with serum levels of C4A, and the number of short genes with serum levels of C4B. Thus, the polygenic and gene size variations of C4A and C4B contribute to the quantitative traits of C4 with a wide range of serum protein levels and hemolytic activities, and consequently the power of the innate defense system.
2.2 INTRODUCTION

Diversities in immune functions are mainly achieved by nucleotide polymorphisms / mutations and somatic recombinations of gene segments (Litman et al 1999, Wagner and Neuberger 1996). Human complement component C4 represents a new paradigm of complex, inborn immune diversity (Yu et al 2003). Phenotypically there are acidic C4A and basic C4B proteins with multiple polymorphic variants in each class (Mauff et al 1998). The plasma or serum protein levels of total C4 vary between 100 and 1000 mg/L among different individuals, as do the relative quantities of C4A and C4B proteins. Although this phenomenon was observed over a decade ago (Averill and Bernal 1984, Chrispeels et al 1989, Hammond et al 1992, Holme et al 1988, Moulds et al 1990, O'Neill and DuPont 1979, Rebmann et al 1992, Teisberg et al 1988, Uko et al 1986, Welch et al 1985), the molecular genetic basis for the qualitative and quantitative variations of human complement C4 was often accounted for inaccurately. Until recently, the sophistication of human C4 genetics has not been fully appreciated. A two-locus, C4A-C4B model with null alleles of either locus was used to explain the great variation in the levels of C4A and C4B proteins in a population (Awdeh and Alper 1980, O'Neill et al 1978a). There are actually a frequent, dichotomous gene size variation, polygenic and modular duplications of C4A and C4B together with their flanking genes RP1 or RP2, CYP21A or CYP21B, and TNXA or TNXB in the major histocompatibility complex (MHC) (Figure 2.1). Existing at the germline levels and inherent to different individuals, there are monomodular, bimodular, trimodular or quadrimodular structures of discrete segments in the MHC containing one, two, three or four RP-C4-CYP21-TNX (RCCX) modules (Blanchong et al 2000, Chung et al 2002b, Yang et al 1999). Each C4 gene may
code for C4A or C4B. Each C4A or C4B gene may be 21 kb (long) or 14.6 kb (short) in size (Yu 1991). The long C4 gene contains a 6.36-kb endogenous retrovirus HERV-K(C4) in its intron 9 (Dangel et al 1994, Schneider et al 2001). Recently, we showed that the bimodular RCCX haplotypes constitute slightly more than two-thirds of the Caucasian population in the mid-western United States, while the frequencies of monomodular and trimodular RCCX haplotypes were 0.17 and 0.14, respectively (Blanchong et al 2000). Quadrimodular RCCX haplotypes with four C4 genes in a row are rare in Caucasians, but relatively common in Asians (Chung et al 2002a, Chung et al 2002b). The impacts of the C4 polygenic and gene size variations on C4 protein levels and functional activities have not been accurately studied in a population.

Complement component C4 is important in immunity, tolerance and autoimmunity (Carroll 1998b, Porter 1983, Reid and Porter 1981, Walport 2001a). It plays essential roles in linking the recognition pathways of the complement system initiated by Ag-Ab or Ag-mannan-binding lectin complexes to the effectors of the humoral immune response. Complete deficiency of complement C4 (i.e., both C4A and C4B) is one of the most penetrant genetic risk factors in autoimmune disease such as systemic lupus erythematosus (SLE) (Atkinson and Schifferli 1999, Yang et al 2004). Complete or partial deficiencies of C4B are related to vulnerability and severity of microbial infections (Bishof et al 1990, Jaatinen et al 1999). Complete deficiencies of C4 in humans or guinea pigs have impairments in the secondary immune response and switching of IgM to IgG (Bottger et al 1986, Ellman et al 1971, Finco et al 1992). In an immunization process, it has been shown that low concentrations of C4 in mice (Milich 2002) or partial deficiencies of C4A in humans are associated with higher frequencies of

Therefore, an accurate account of the phenotypic and genotypic diversities of C4A and C4B in a human population both qualitatively and quantitatively is critically important for epidemiological studies of disease associations with C4.

In this study, we report a comprehensive investigation of human C4A and C4B serum protein levels and functional hemolytic activities in a central European population with precisely defined status for C4A and C4B gene dosage, gene size, and RCCX modules. The study reveals higher C4 serum protein levels and hemolytic activities among individuals with short C4 genes than those with long C4 genes only. It also demonstrates that human serum C4 protein levels were determined by the dosage and size of C4 genes and, unexpectedly, the body mass index (BMI) of an individual.
2.3 MATERIALS AND METHODS

2.3.1 Study population

The study was performed in 128 healthy Hungarian subjects (40 males, 88 females). The subjects were 44.7 ± 10 (mean ± SD) years old. Their BMI were 25.5 ± 4.7 kg/m², and serum total cholesterol and triglyceride concentrations were 5.48 ± 1.02 mM, and 1.44 ± 0.90 mM, respectively. These individuals participated in a regular medical survey and gave their informed consent for the use of their serum/plasma and DNA samples for the present study. For ethical reasons, after their computer registration, the data were unlinked from the subjects so that their identities could not be traced. The study was approved by the Ethical Committee of the Semmelweis University (Budapest, Hungary). Peripheral blood samples were obtained by venipuncture and processed immediately. EDTA-plasma and serum samples were aliquoted and stored at −80°C and thawed only immediately before the assays/measurements were performed. Genomic DNA samples were stored at 4°C.

2.3.2 Measurement of serum C4 and C3 concentrations

Concentrations of C4 and C3 were determined immunochemically by single radial immunodiffusion (Kohler and Müller-Eberhardt 1967), using commercial antibodies against C3 and C4 (DiaSorin, Stillwater, MN). Human Serum Protein Calibrator (DAKO A/S, Glostrup, Denmark) was used as standard.

2.3.3 Measurement of the complement C4HA

C4 hemolytic activities (C4HA) were measured by the effective molecule titration method (Nelson et al 1966). Briefly, different dilutions of the serum samples were
incubated with EAC_{1gp} at 30°C for 20 min; purified guinea pig C2 was added and further incubated at 30°C for 15 min. Finally, C-EDTA (guinea pig serum diluted in buffer containing 10 mM EDTA) was added and incubated at 37°C for 60 min. After centrifugation, OD values were read at 412 nm. C4HA in serum was expressed in complement hemolytic units CH_{63} U/ml, which is the amount of complement component required to lyse 63% of the indicator cells in standard C4 assays. Normal values of the C4HA in the laboratory were between 5,000 and 70,000 CH_{63} U/ml. Hemolytic titration of C4 was standardized using aliquots stored at –70°C of the same normal human serum pool at each measurement.

2.3.4 Preparation of genomic DNA, Southern blot analysis, and DNA probes

Genomic DNAs were prepared from peripheral blood mononuclear cells (Miller et al 1988). Southern blot analyses of TaqI-digested or PshAI-digested genomic DNAs were performed, as described in previous publications (Blanchong et al 2000, Yu et al 2002). The TaqI genomic blots were hybridized with three specific probes corresponding to RP, CYP21 and TNX. The results revealed the presence and dosage of RP1-C4L (7.0 kb), RP1-C4S (6.4 kb), RP2-C4L (6.0 kb) and RP2-C4S (5.4 kb); the presence and relative dosage of CYP21B (3.7 kb) and CYP21A (3.2 kb); and the presence and relative dosage of TNXB (2.5 kb) and TNXA (2.4 kb). The number of RCCX modules was independently confirmed by PshAI-RFLP using a RP 3' probe to determine the relative dosage of RP1 (7.2 kb) and RP2 (8.3 kb) fragments. The relative dosages of C4A and C4B genes were determined by PshAI RFLP using a C4d specific probe corresponding to C4 gene exons 28-31 (Figure 2.1).
2.3.5 Determination of C4 protein polymorphism by immunfixation and immunoblot analysis

EDTA-plasma samples were digested overnight with neuraminidase at 4°C, followed by carboxypeptidase B digest for 30 min at room temperature, and resolved with high voltage agarose gel electrophoresis (HVAGE). C4 proteins were immunofixed with goat anti-human C4 sera. Gels were blotted to remove diffusible proteins, and the immune complexes with C4 proteins were stained with Easy-Stain (Invitrogen, Carlsbad CA), as described previously (Awdeh et al 1979, Sim and Cross 1986, Yu et al 2002). The relative band intensities of C4A and C4B allotypes in each sample were quantified by densitometry, using an Epson Expression 1600 Scanner, and analyzed by ImageQuant Version 5.0 software. Ambiguous C4A or C4B allotypes were further investigated with immunoblot analysis of HVAGE-resolved gels using anti-Rg1 and anti-Ch1 monoclonals provided by the VIIth Complement Genetics Workshop (Mauff et al 1998).

2.3.6 Statistical analysis

Parametric tests (two-sample t test, paired t test, Pearson correlation analysis) were used for the normally distributed (C3 and C4 levels, C4HA, BMI) data. All tests were two-tailed. Multiple linear regression was used to evaluate potential confounders. Statistical analysis was performed by GraphPad Prism 3.0 (GraphPad Software, San Diego, CA, www.grphpad.com) and SPSS 10.0 (SPSS, Chicago, IL) software.

2.3.7 Additional declarations

This is a collaborative study with a group of scientists in Budapest, Hungary. Healthy Caucasians originally participating in a regular medical survey were recruited
from Semmelweis Medical University, Budapest, Hungary. After the withdrawal of peripheral blood samples, EDTA-plasma, serum and genomic DNA were isolated locally. EDTA-plasma and genomic DNA samples from each individual were sent to our laboratory for immunofixation, immunoblot and southern blot analyses. Serum sample was used by our collaborators in Hungary for measurement of the complement C4HA and C4 and C3 concentrations.

2.4 RESULTS

2.4.1 Genotypic variation of human complement components C4A and C4B

The study population consisted of 128 healthy individuals (40 males and 88 females) from Budapest, Hungary. Molecular genetic approaches were applied to determine the genotypic diversities of the constituents of RCCX modules. The total number of C4 genes present in a diploid genome (i.e., gene dosage) was elucidated by TaqI RFLP that also yielded information on the details of the RCCX variants such as the presence and relative dosage of RP1 and RP2 linked to a long or a short C4 gene, and presence and relative dosage of CYP21A and CYP21B, and TNXA and TNXB (Figure 2.1B). The TaqI RFLP results were independently confirmed by PshAI RFLP analysis for the relative dosage of RP1 and RP2. The PshAI Southern blots were also used to determine the relative dosage of C4A and C4B using a specific C4d genomic probe, as one of the five nucleotide changes specific for C4A and C4B created a new PshAI cleavage site in C4A (Chung et al 2002c, Yu et al 1986). The polymorphism of C4A and C4B proteins were elucidated by HVAGE of EDTA-plasma, followed by immunofixations. Selected plasma samples with ambiguous C4A or C4B allotypes were further analyzed by immunoblot analysis using anti-Rg1 and anti-Ch1 monoclonals.

Demonstration of variations in the RCCX constituents: Figure 2.2 illustrates the genotypic characterization of the RCCX constituents and phenotypic polymorphism of C4 in 10 individuals. Eight subjects had bimodular RCCX structures from both haplotypes; among them were homozygous LL/LL (L, long) (MK1, 11 and 13), heterozygous LL/LS (S, short) (MK2, 3 and 5) and homozygous LS/LS (MK12). Of these eight individuals who each had a total of four C4 genes in a diploid genome, five had two
C4A and two C4B genes (MK2, 3, 8, 12 and 13; Figure 2.2C). Two subjects, MK1 and MK5, had three C4A and one C4B. One subject (MK11) had four C4A, but no C4B. The remaining two subjects had heterozygous bimodular and monomodular RCCX structures, i.e., LS/S in MK6 and LS/L in MK9. Both of them had one C4A and two C4B genes.

Results of the immunofixation experiment for C4A and C4B protein allotypes are shown in Panel D of Figure 2.2D. In addition to C4A3 that was observed in every individual, C4A2 was present in MK2 and C4A6 was present in MK3. C4B1 was detectable in all except MK5 who had C4B2, and MK11 who was deficient in C4B. The protein band intensities of C4A and C4B were consistent with the C4A and C4B gene dosage results. For examples, MK1 had three C4A and one C4B genes (Figure 2.2C) and the protein band intensities of C4A3 allotype were about 3 times greater than that of C4B1 (Figure 2.2D); MK11 had complete absence of the C4B gene (Figure 2.2C) and that was matched with the absence of C4B protein (Figure 2.2D).

The TaqI and PshAI RFLP patterns for the RCCX modules revealed a common pattern for bimodular structures with RP1-C4-CYP21A-TNXA-RP2-C4-CYP21B-TNXB, and a common pattern for monomodular structures with RP1-C4-CYP21B-TNXB (Figure. 1.1B). A single anomaly was observed in MK6, whose markers for CYP21A and TNXA in the TaqI RFLP appeared to be absent, although he had a bimodular LS and a monomodular S structure. The PshAI RFLP hybridized to RP 3' probe revealed the presence of a 3.5-kb, rearranged fragment that is characteristic of RP2-TNXA+120. In other words, MK6 has an unusual LS haplotype that is characterized by the presence of two CYP21B genes and a rearranged TNXA with the 120-bp sequence that is normally
present in exon 36-intron 36 of the TNXB gene (i.e., RP1-C4-CYP21B-TNXA+120 - RP2-C4-CYP21B-TNXB) (Rupert et al 1999).

The RCCX and complement C4 polygenic variations in the Budapest population: The frequencies of RCCX length variants, namely, monomodular L and S; bimodular LL and LS; and trimodular LLL, LSS and LSL, are listed in Table 2.1. The most common RCCX haplotype in the Hungarian population was the bimodular LL with a frequency of 0.500. Another bimodular haplotype LS had a frequency of 0.289. The frequencies for monomodular L and S haplotypes were 0.047 and 0.086, respectively. The frequencies of the trimodular haplotypes LLL, LSS and LSL were 0.039, 0.027 and 0.004, respectively. Altogether, the frequencies of monomodular, bimodular and trimodular RCCX haplotypes were 0.133, 0.797 and 0.070, respectively.

In this central European population, the number of C4 genes in a diploid genome varied from two to five. The majority of individuals had four C4 genes per genome, which had a frequency of 0.656. Individuals with C4 gene dosages of three and five had frequencies of 0.211 and 0.117, respectively. Only two individuals had a dosage of two C4 genes (frequency: 0.016). We did not detect any individuals in the study population with six C4 genes.

In the 128 genotyped individuals, there were 496 C4 genes. Among these genes, 77.4% belonged to the long form with HERV-K(C4) integrated into the ninth intron; 22.6% belonged to the short form without the endogenous retrovirus in the C4 gene. The average C4 gene number in the normal Hungarian population (gene index) was 3.875, of which C4A had a gene index of 2.0 and C4B, 1.875. The frequency of individuals with gene dosage of two C4A was 0.570, and with two C4B was 0.700. Overall, one-fifth of
the study population had a single C4A gene in a diploid genome; the same frequency was also found in subjects with single C4B genes. The frequency of individuals with a gene dosage of three C4A was 0.190, and that of three C4B was 0.080. Three individuals (frequency: 0.023) had four C4A genes, and one single individual (frequency: 0.008) had four C4B genes in a genome. Two subjects had no C4A gene (frequency: 0.016), and one individual had no C4B gene (frequency: 0.008).

2.4.2 Qualitative variations of C4A and C4B

The polymorphism of C4A and C4B proteins was investigated by immunofixation of EDTA-plasma, resolved by HVAGE. Samples showing ambiguous or unusual C4A and C4B allotypes were further analyzed by immunoblot analysis using anti-Rg1 and anti-Ch1 mAbs. The frequencies of C4A and C4B allotypes detected in this Hungarian study population are listed in Table 2.2.

Among the 256 C4A genes in the study population, 249 coded for a detectable protein. The most common C4A allotype was A3, which had a frequency of 85.5%. C4A2, and A6 had frequencies of 7.81 and 1.95%, respectively. Rare C4A allotypes detected in the study population included A92, A12, A4 and A5, which had a combined frequency of 1.95%. In the study population, there were seven suspected C4A mutant genes not producing C4A proteins. Among them, three (1.17%) were confirmed by sequence-specific primer PCR to have nonsense mutations caused by a 2-bp insertion at exon 29 (Barba et al 1993, Rupert et al 2002).

In the study population, all 240 C4B genes coded for C4B proteins, of which C4B1 had a frequency of 83.8%. C4B2 and B92 had frequencies of 10.8% and 1.25%,
respectively. Rare C4B allotypes detected include B94, B12, B15, B3, B4, B5 and B6, which had a combined frequency of 4.17%.

2.4.3 Quantitative diversities of C4A and C4B proteins

The C4 protein level in each individual was measured by single radial immunodiffusion of serum proteins. In the protein-allotyping gels, the band intensities for C4A and C4B were quantified by densitometry and the corresponding C4A and C4B isotype protein levels calculated from the total serum C4 concentrations. The relationship of total C4, C4A and C4B serum concentrations with respect to the dosages of total C4 (C4A plus C4B), C4A, C4B, long C4 genes and short C4 genes was analyzed and shown in Figure 2.3. For total C4 gene dosage, the population was categorized into three groups: individuals with 4 C4 genes (medium), < 4 C4 genes (2 or 3; low), and >4 C4 genes (5; high). With respect to C4A gene dosage, the population was grouped into individuals with <2 C4A genes (0 or 1; low), and ≥2 C4A genes (2, 3 or 4; high). The same categorization was applied for C4B genes. With respect to long C4 genes, the population was divided into two groups: 0-2 long C4 genes, and 3-5 long C4 genes. With respect to short C4 genes, the population was divided into two groups: with (i.e., 1-3) and without (i.e., 0) short C4 genes.

Total serum C4 protein concentrations and C4 gene dosages: As shown in Figure 2.3A, there was a positive correlation of serum C4 concentrations with total C4 gene dosages. The low C4 gene dosage group (gene dosages: 2-3; n = 29) had a mean C4 concentration of 0.33 ± 0.10 g/L, which was significantly lower than that of the medium gene dosage group (gene dosage: 4; n = 84; concentration 0.46 ± 0.15 g/L; p < 0.001), and the high gene dosage group (gene dosage: 5; n = 15; concentration 0.50 ± 0.16 g/L; p
The total serum C4 concentration was significantly higher in individuals with 2, 3 or 4 C4A genes (concentration 0.46 ± 0.16 g/L; n = 100; p = 0.0196) than in those with 0 or 1 C4A gene (concentration 0.38 ± 0.13 g/L; n = 28) (Figure 2.3B). With respect to the number of the C4B genes, however, there was no relevant correlation with the total C4 serum concentration (Figure 2.3C). With respect to the dosage of the long C4 genes, there was also no relevant difference in the C4 serum concentrations between the low dosage and the high dosage groups (Figure 2.3D). On the contrary, a higher total C4 serum concentration was observed in the population group with one or more short C4 genes (concentration 0.47 ± 0.17 g/L; n = 80) than the group without any short C4 genes (concentration 0.39 ± 0.11 g/L; n = 48; p = 0.0095; Figure 2.3E).

Serum C4A and C4B concentrations and C4 gene dosages: We then analyzed the impact of gene dosages for total C4, C4A, C4B, long genes and short genes on the serum protein concentrations of C4A and C4B. The mean C4A serum concentration in the low (i.e., < 2), medium (i.e. = 2) and high (i.e., > 2) C4 gene dosage groups were 0.13 ± 0.05, 0.23 ± 0.08, and 0.26 ± 0.08 g/L, respectively. There was a close, positive correlation of C4A serum concentrations with the total C4 gene dosage (p < 0.001; Fig 2.3F). As expected, the serum C4A concentration was a function of C4A gene dosage (p < 0.0001; Figure 2.3G), and was weakly, but inversely related to the C4B gene dosage (Figure 2.3H). With respect to C4 gene size, higher serum C4A concentrations were present in individuals with high dosage (concentration 0.23 ± 0.08 g/L) than those with low dosage of long C4 genes (concentration 0.18 ± 0.09 g/L; p = 0.0028; Figure 2.3I). The presence of short C4 genes, however, did not appear to have a correlation with the C4A protein concentration (p = 0.9392; Figure 2.3J).
The mean C4B serum concentrations were positively correlated with C4B gene dosage (p < 0.0001; Figure 2.3M), and a highly significant correlation was noted with the presence of one or more short C4 genes (p < 0.0001; Figure 2.3O). The serum C4B level was not related to C4A gene dosage (Figure 2.3L). There was a mild but significant inverse correlation of serum C4B levels with the number of long C4 genes (p = 0.048; Figure 2.3N). Unlike the case for C4A, there was no significant correlation of C4B protein concentrations with total C4 gene dosages (Figure 2.3K).

2.4.4 Hemolytic activities, C4 gene dosages, C4 gene sizes and serum C4 concentrations

C4HA of the serum samples were determined by effective molecule titration and expressed in CH₆₃ U/ml. The C4HA from each sample was scatter plotted against the C4 gene dosages (Figure 2.4). There was a direct correlation of the mean C4HA with an increase of total C4 gene dosage. This was most obvious when the comparison was made between the high C4 gene dosage group and the low gene dosage group (p < 0.01; Figure 2.4A), and also between the groups with one to three short C4 genes and without short C4 genes (p = 0.0063; Figure 2.4E). However, there was no clear-cut correlation of C4HA with the gene dosage of C4B (Figure 2.4C) or with the dosage of long C4 genes (Figure 2.4D).

When the relations of serum C4 concentrations and C4HA were examined, it was found that C4HA strongly correlated to the total C4 serum concentrations (Figure 2.4F), and to the serum levels of C4A (Figure 2.4G) and C4B (Figure 2.4H) (p < 0.0001). The correlation coefficients were ~0.5. The strengths of correlation between C4HA and C4A protein levels, and between C4HA and C4B protein levels were similar.
To further illustrate the intricate relations among the size of C4 genes, C4 serum protein concentrations, and hemolytic activities, we focused on the most common gene dosage group, i.e., subjects with four C4 genes (two C4A and two C4B) and organized in one of the following three RCCX configurations: LL/LL, LL/LS and LS/LS. In comparison with LL/LL homozygotes, individuals with LS/LS structures had higher serum C4 protein concentrations (p < 0.05; Figure 2.5A) and higher C4 hemolytic activities (p < 0.01; Figure 2.5B). Although not reaching a statistically significant level, heterozygous LL/LS had values between LL/LL and LS/LS for both serum C4 protein concentrations and C4HA. Compared with those having homozygous LL/LL, the C4A and C4B serum protein levels increased by 26.0% and 34.0%, respectively, in LL/LS heterozygotes, and by 34.0% and 40.0% in LS/LS homozygotes. The increase of C4B protein levels was significant in both LL/LS (p < 0.01) and LS/LS (p < 0.01). This phenomenon probably suggested that the presence of short C4 gene(s) at the second locus not only significantly increased the expression of the C4B proteins, but also had a considerable, positive impact on the expression of C4A proteins. It appeared that the presence of HERV-K(C4) reduced the expression of serum C4 and the C4 hemolytic activities in a dose-dependent manner.

2.4.5 BMI and serum C4 and C3 concentrations

In view of the emerging roles of several complement proteins in lipid metabolism and the strong correlation of serum C3 levels and BMI in healthy people, we investigated whether there were positive relationships between BMI and C4 gene dosage; total C4, C4A, C4B serum protein concentrations; and C4HA (Figure 2.6). BMI was not related to the C4 gene dosage or C4HA (Figure 2.6, A and E, respectively). However, there was a
modest, but significantly positive correlation between BMI and serum protein concentrations of C4, C4A and C4B (Figure 2.6, B, C and D, respectively).

As expected, C3 serum levels were found to be strongly correlated to the BMI (Figure 2.6F). Approximately the same coefficients were found in the males and females for each correlation depicted in Figure 2.6. The only exception was the correlation between serum C3 concentrations and BMI. In males, we found a very strong correlation ($r = 0.675, p = 0.0004$), whereas in females the correlation was weaker ($r = 0.346, p = 0.0041$).

The serum C4 concentration did not correlate to the age of subjects or the serum total cholesterol levels (correlation coefficient $r = 0.091, p = 0.316$) and triglyceride ($r = 0.082, p = 0.362$). By contrast, there was a significant positive correlation between C3 levels and the total cholesterol concentration ($r = 0.246, p = 0.015$), and a very strong correlation between C3 levels and concentrations of triglycerides ($r = 0.510, p < 0.001$).

Using multiple linear regression analysis, we found the correlation between BMI and serum C4 concentrations to be independent from that between BMI and serum C3 concentrations (Table 2.3). In the whole population, a strong correlation between C3 and BMI was found when no adjustment was applied. However, when serum cholesterol and triglyceride levels were adjusted to the age, gender, cholesterol and triglyceride levels, the correlation coefficient between C3 concentration and BMI became marginally significant. This difference became more pronounced when the analysis was restricted to the subjects with only bimodular RCCX modules. In this group, no significant correlation ($p = 0.356$) was noted between C3 levels and BMI after adjustment. Similar disappearance of the correlation between C3 and BMI ($p = 0.318$) after adjustment was
noted in subjects with at least four RCCX modules (i.e., those with monomodular/monomodular and monomodular/bimodular RCCX combinations were excluded). By contrast, adjustments to age, gender and lipid levels did not negatively affect the strength of the correlation between C4 levels and BMI and in the group of subjects with at least four RCCX modules: the adjusted correlation between C4 concentration and BMI became highly significant instead (p = 0.009) (Table 2.3).
2.5 DISCUSSION

In this study, we report three important novel findings. First, we demonstrated that the serum concentrations of C4A and C4B proteins and total amounts of C4 proteins were determined by the number of long and short C4 genes. Second, in individuals with medium or high C4 gene dosages (i.e., $\geq 4$ genes), a significant positive correlation was found between BMI and C4 serum levels, which in several aspects was quite different from the correlation between BMI and serum C3 concentrations. Third, by using precise and definitive methods in genotyping and phenotyping experiments, we demonstrated the correlation between C4 gene dosage and C4 proteins concentrations. The results provide a foundation for the differences on an important factor of innate immune response. They enable a better understanding of the genetic basis for the qualitative and quantitative diversities of human complement C4, which include variations in immunochemical and functional properties of C4A and C4B, and a wide range of serum protein levels and hemolytic activities.

2.5.1 Factors determining the serum C4 protein levels

Under pathological conditions, low serum or plasma C4 levels may be caused by excessive consumption due to complement activation as a result of microbial infections or abnormalities such as a deficiency of the complement C1 inhibitor (Frank et al 1976, Gompels et al 2002), and the presence of nephritic factors or autoantibodies that stabilizes the C1 complex (He and Lin 1998, Trouw et al 2002). Therefore, the serum concentrations of total C4 and C3 and their split products have been used as indicators for disease activities of autoimmune disease, and inflammatory responses including graft rejections (Ho et al 2001, Moulds et al 1993, Wild et al 1990). However, there is a more
fundamental and sophisticated genetic mechanism leading to a wide range of serum protein levels of total C4, C4A and C4B in the general population. Besides the polygenic variation from two to five genes among different individuals in our study population, the gene size dichotomy adds an additional dimension to the genetic complexity of human C4. Close to three-quarters (77.4%) of the C4 genes in this study population were long and about one-quarter (22.6%) of the C4 genes were short. This intrinsic diversity in the C4 gene number and gene size is extraordinary in human genetics, which we suggest to be an adaptation of the innate immune system to provide a wide spectrum of immune efficiencies and strengths to defend against parasites. We have established the sophistication of the C4 genetics in a Caucasian population in mid-western region of the United States (Blanchong et al 2000). This current study does not only support and validate results of the Ohioan study, it also provides an essential link between the C4 genetic diversity with the phenotypic and functional variations. The C4 gene number and gene size and, unexpectedly, the BMI are all relevant in determining the C4 serum protein concentrations, which impact the power of the complement system as reflected by the hemolytic assays. In a separate report, we have delineated in the same Caucasian populations the complex relationship between complement C4 and RCCX modular variations with two common single nucleotide polymorphisms of the TNF gene TNFA, −238 G/A and −308 G/A (Vatay et al 2003).

The serum C4A protein levels were closely associated with C4A gene dosage, the number of long C4 genes, and the total C4 gene dosages. This phenomenon may reflect the fact that most C4A genes are located in the first RCCX module that frequently contains the endogenous retrovirus HERV-K(C4), and an increase in C4 gene dosage has
a higher tendency to increase the number of C4A genes. In contrast, serum C4B protein levels were mainly the functions of gene dosages of C4B and short C4 genes. The latter reflects the fact that more short C4 genes code for C4B proteins than for C4A proteins.

2.5.2 The significance and implication of the long and short C4 genes

It is of interest to note that individuals containing one or more short C4 genes consistently have higher serum total C4 and C4B concentrations than those with long C4 genes only. Higher expression levels of serum C4B proteins than C4A proteins in individuals with equal number of C4A and C4B genes had been observed previously in specific haplotypes such as HLA B18 DR2 C4A4 B2 (Truedsson et al 1989, Truedsson et al 1995a). This B18 DR2 haplotype has a bimodular RCCX structure LS, and it is plausible that the short C4B gene is responsible for the higher serum level of the C4B2 protein than the C4A4 protein. However, the positive impact of short C4 genes seemed more far reaching because it affected not only the C4B (and the total C4 protein levels), but also the levels of C4A proteins, albeit to a lesser extent. This was best illustrated in subjects with bimodular RCCX structures from both copies of chromosome 6 and contained two C4A and two C4B genes. Individuals with homozygous LS/LS had mean serum C4A and C4B protein levels 34.0% and 40.0% higher than those from homozygous LL/LL, respectively. The levels of C4A and C4B proteins from individuals with heterozygous LL/LS were higher than those with LL/LL but lower than those with LS/LS (Figure 2.5). It is likely that the 6.36-kb endogenous retrovirus in the long C4 genes retarded the gene transcription rate because of the larger size of the heteronuclear transcripts, and consequently reduced the corresponding protein products. Consistent
with this notion, we found that individuals with LSS/LL have serum C4 protein levels 41.0% higher than those with LLL/LL.

The selection advantage of having an endogenous retrovirus in three-quarters of C4 genes in the Caucasian population is a matter of speculation. It was proposed that the reversely oriented HERV-K(C4) in the intron 9 of the long C4 gene would lead to the production of an antisense transcript of the ancient retrovirus whenever the long C4 gene is transcribed (Dangel et al 1994). This antisense retroviral RNA transcript with multiple defective mutations could confer the host selection advantage in the defense against retroviral infections. This is because partial hybridizations of the antisense endogenous retroviral RNA transcripts to the proviral RNA genomes would trigger the cellular RNase surveillance and the IFN systems to degrade the double-stranded retroviral hybrids with multiple mismatches. Thus, the host endogenous retroviruses could be a form of innate defense system against retroviral infections (Dangel et al 1994, Schneider et al 2001).

2.5.3 C4HA as a measure of the strength of the complement system

Results of our experiments on the functional activities of C4A and C4B using a conventional hemolytic assay revealed a direct correlation of the hemolytic activities with serum C4 protein concentrations. Higher C4 gene dosage and presence of short C4 gene(s) increased the serum C4 protein levels and hemolytic reactivities. Importantly, this observation reflects the influence of the gene size and polygenic variations of human C4 on the intrinsic strength of the complement system.

In in vitro studies using purified human C4 proteins in most heterologous assays using rabbit Abs and C4-deficient guinea pig complement proteins, activated human C4A proteins have high binding affinities to amino group-containing Ags or immune
complexes; activated C4B proteins have an internal catalytic mechanism that favors rapid covalent binding to hydroxyl group-containing Ags through a transesterification mechanism (Dodds et al 1996, Isenman and Young 1984). Unexpectedly, the C4 hemolytic activities from 128 samples employing conventional hemolytic assays were indifferent to the C4A and C4B isotype composition in the serum. In other words, serum C4A and C4B proteins manifested similar activities in the fluid-phase hemolytic assays, if their serum protein levels were the same. Indeed, differential reactivities between C4A and C4B for bindings to immune complexes, human or sheep erythrocytes and complement receptor CR1, and for prevention of immunoprecipitation and solubilization of immune aggregates could be demonstrated only when the C4A and C4B proteins were purified by an immunochemical method or resolved by gel electrophoresis (Awdeh and Alper 1980, Gatenby et al 1990, Gibb et al 1993, Paul et al 1988, Reilly 1999, Reilly and Mold 1997, Schifferli et al 1987). A differential reactivity between C4A and C4B could not be detected in assay systems when whole human sera were used as a source of C4A or C4B proteins or in reconstituted assays when human C4A- or C4B-deficient sera or human sera with selectively depleted C4 were used (Holme et al 1992, Klint et al 1995, Reilly 1999, Varga et al 1991). One explanation is that the native or activated C4A and/or C4B proteins are regulated or protected by protein(s) in the autologous sera, which mask their differential chemical reactivities in a conventional assay system.

2.5.4 Correlation between BMI and serum concentrations of C4

The observation of a correlation between human serum C4 levels and BMI is novel. When calculated by multiple linear regression analysis adjusted to sex, age and serum lipid levels, the correlation of C4 and BMI was stronger in the subjects with bimodular
RCCX modules only than in the entire study population, and it was the strongest when subjects with monomodular / monomodular and monomodular / bimodular combinations were excluded from the analysis (Table 2.3). The present findings indicate that the mechanism of correlation between C4 and BMI is independent and different from that of complement C3 and BMI. In contrast to C3, there is no correlation between serum lipid levels and serum C4 concentrations. Moreover, at multiple linear regression analysis, C3 and BMI did not significantly correlate after adjustment to serum lipid levels; the same adjustment did not negatively influence the correlation between C4 serum levels and BMI. The correlation between C3 and BMI was stronger in the males than in the females, whereas no gender-related difference could be detected in the relationship of C4 levels and BMI. Why the serum C4 levels are higher in obese than in lean subjects remains to be investigated.

In conclusion, we have shown the C4 polygenic and gene size variations contributing to a wide range of serum levels for C4A and C4B. This phenomenon is relevant. When a patient is diagnosed and treated for a disease, the information on the number of C4A and C4B genes and their serum levels may be helpful. This is because the C4 data would reflect the strength of innate immunity, and also the vulnerability to infections and susceptibility to autoimmune diseases. It is important to note that the serum C4A and C4B are predominantly secreted by the liver and the results presented in this work largely reflected how the C4 gene size and polygenic variations and the BMI influenced the production of C4 by the liver. The influence of C4 polygenic and gene size variations on the regulation mechanisms of C4A and
C4B protein biosynthesis in local tissues such as the kidneys, heart, CNS, thyroids and adrenals is yet to be determined.
Figure 2.1 The gene size and polygenic variations of human complement C4 and RP-C4-CYP21-TNX (RCCX modules) in the MHC class III region. A, A molecular map of the genes in the MHC complement gene cluster. Horizontal arrows represent gene configurations; filled, vertical arrows represent locations to where the DNA probes hybridized in a TaqI genomic Southern blot analysis to determine the RCCX modular variations; gray, vertical arrows represent locations to where the DNA probes hybridized in a PshAI genomic Southern blot analysis to determine the ratio of C4A and C4B genes. B, Common RCCX length variants in healthy Caucasians. Characteristic TaqI restriction fragments of each RCCX haplotype in a Southern blot analysis are shown on the right.
Figure 2.2 Genotypic variations of RCCX modules, C4A and C4B gene dosages, and protein polymorphisms. A, TaqI RFLP to show RCCX modular variations. Hybridization probes include 3’ RP cDNA, CYP21 genomic, and 3’ TNX genomic fragments. B, PshAI-RFLP to show number of RCCX modules by determining the relative dosages of RP1 and RP2. A 3’ RP cDNA probe was used for hybridization. C, PshAI-RFLP to show relative dosages of C4A and C4B genes. Hybridization probe used was a genomic fragment corresponding to C4 exons 28–31. D, Allotyping of human C4A and C4B proteins. EDTA plasma C4 proteins were resolved by HVAGE, processed by immunofixation using goat polyclonal antisera against human C4. Notice the presence of restriction fragments corresponding to CYP21B and TNXB only (marked by arrows, A), and the presence of a novel 3.5-kb PshAI fragment corresponding to a rearranged TNXA-XB recombinant (B) in subject MK6. Subject MK11 was completely C4B deficient, as there were no C4B gene (marked by an arrow, C) and no C4B protein (marked by an arrow, D).
Figure 2.3 Analyses of C4 serum concentrations with C4 gene dosage and gene size variations. A–E, Show the relationships of total serum C4 levels with C4 polygenic and gene size variations. F–J, Show the relationships of serum C4A proteins with C4 polygenic and gene size variations. K–O, Show the relationships of serum C4B proteins with C4 polygenic and gene size variations. A, F, and K, The p values for Tukey post hoc test of one-way ANOVA are indicated. For all other panels, p values for unpaired t test are indicated.
Figure 2.4 Analyses of C4 hemolytic activities with C4 polygenic and gene size variations (A–E), and with serum C4 concentrations (F–H). Serum hemolytic activity of C4 expressed in CH50 U/ml in the sera of 128 healthy subjects with low (gene dosages = 2–3), medium (gene dosage = 4), or high (gene dosage = 5) dosages of C4 genes (A); low (gene dosage = 0 or 1) or high (gene dosages = 2–4) number of C4A genes (B); or C4B genes (C), low (gene dosages = 0–2) or high (gene dosages = 3 to 5) number of long (L) C4 genes (D), and none or 1–3 short (S) C4 genes (E). A, The p values for Tukey post hoc test of one-way ANOVA are indicated. B–E, The p values for unpaired t test are indicated. F–H, The Pearson correlation coefficients and their significance are indicated.
Figure 2.5 The impact of long and short C4 genes on the serum C4 protein concentrations and C4HA. Eighty healthy subjects with bimodular/bimodular RCCX haplotypes corresponding to LL/LL, LL/LS, or LS/LS configurations were chosen for analysis. The p values for Tukey post hoc test of one-way ANOVA are indicated.
Figure 2.6 Analyses of BMI and C4 polygenic variations (A), serum C4 concentrations (B–D), C4HA (E), and serum C3 concentration (F). A, The p values for Tukey post hoc test of one-way ANOVA are indicated. B–F, The Pearson correlation coefficients and their significance are indicated. BMI was expressed in kg/m².
<table>
<thead>
<tr>
<th></th>
<th>Hungary</th>
<th>Ohio *</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Frequency</td>
</tr>
<tr>
<td>1. RCCX Length Variants</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>22</td>
<td>0.086</td>
</tr>
<tr>
<td>L</td>
<td>12</td>
<td>0.047</td>
</tr>
<tr>
<td>LL</td>
<td>128</td>
<td>0.500</td>
</tr>
<tr>
<td>LS</td>
<td>74</td>
<td>0.289</td>
</tr>
<tr>
<td>LLL</td>
<td>10</td>
<td>0.039</td>
</tr>
<tr>
<td>LSL</td>
<td>1</td>
<td>0.004</td>
</tr>
<tr>
<td>LSS</td>
<td>7</td>
<td>0.0273</td>
</tr>
<tr>
<td>Lx +</td>
<td>2</td>
<td>0.008</td>
</tr>
<tr>
<td>Total</td>
<td>256</td>
<td>1.00</td>
</tr>
<tr>
<td>2. Long (L) and Short (S) C4 genes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L genes</td>
<td>384</td>
<td>0.774</td>
</tr>
<tr>
<td>S genes</td>
<td>112</td>
<td>0.226</td>
</tr>
<tr>
<td>Total</td>
<td>496</td>
<td>1.00</td>
</tr>
<tr>
<td>3. Total Number C4 Genes in a Diploid Genome (C4 Gene Dosage)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.016</td>
</tr>
<tr>
<td>3</td>
<td>27</td>
<td>0.211</td>
</tr>
<tr>
<td>4</td>
<td>84</td>
<td>0.656</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>0.117</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>128</td>
<td>1.00</td>
</tr>
<tr>
<td>Total C4 Gene Index</td>
<td>3.875</td>
<td>3.933</td>
</tr>
</tbody>
</table>

Table 2.1 A summary of RCCX length variants and C4 gene dosages in Caucasians.

(Continued on next page)
<table>
<thead>
<tr>
<th></th>
<th>Hungary</th>
<th></th>
<th>Ohio *</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Frequency</td>
<td>Frequency</td>
</tr>
<tr>
<td>4. C4A Gene Dosage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>0.016</td>
<td>0.007</td>
</tr>
<tr>
<td>1</td>
<td>26</td>
<td>0.203</td>
<td>0.200</td>
</tr>
<tr>
<td>2</td>
<td>73</td>
<td>0.570</td>
<td>0.460</td>
</tr>
<tr>
<td>3</td>
<td>24</td>
<td>0.188</td>
<td>0.295</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>0.023</td>
<td>0.033</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0.007</td>
</tr>
<tr>
<td>Total</td>
<td>128</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>C4A Gene Index</td>
<td></td>
<td>2.0</td>
<td>2.17</td>
</tr>
<tr>
<td>5. C4B Gene Dosage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0.008</td>
<td>0.020</td>
</tr>
<tr>
<td>1</td>
<td>26</td>
<td>0.203</td>
<td>0.320</td>
</tr>
<tr>
<td>2</td>
<td>90</td>
<td>0.703</td>
<td>0.533</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>0.078</td>
<td>0.120</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0.008</td>
<td>0.007</td>
</tr>
<tr>
<td>Total</td>
<td>128</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>C4B Gene Index</td>
<td></td>
<td>1.875</td>
<td>1.77</td>
</tr>
</tbody>
</table>

* taken from Blanchong et al (2000); † undetermined due to a TaqI RFLP at the 5' end of the second C4 gene

Table 2.1 (Continued)
<table>
<thead>
<tr>
<th>Allotype</th>
<th>Frequency (%)</th>
<th>Allotype</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C4A</td>
<td></td>
<td>C4B</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>C4A</td>
<td>all C4*</td>
</tr>
<tr>
<td>A2</td>
<td>20</td>
<td>7.81</td>
<td>4.03</td>
</tr>
<tr>
<td>A3</td>
<td>219</td>
<td>85.5</td>
<td>44.2</td>
</tr>
<tr>
<td>A4</td>
<td>1</td>
<td>0.39</td>
<td>0.20</td>
</tr>
<tr>
<td>A5</td>
<td>2</td>
<td>0.78</td>
<td>0.40</td>
</tr>
<tr>
<td>A6</td>
<td>5</td>
<td>1.95</td>
<td>1.01</td>
</tr>
<tr>
<td>Others</td>
<td>+ 2</td>
<td>0.78</td>
<td>0.4</td>
</tr>
<tr>
<td>AQ0***</td>
<td>7</td>
<td>2.73</td>
<td>1.41</td>
</tr>
<tr>
<td>(2-bp ins)***</td>
<td>(3)</td>
<td>(1.17)</td>
<td>(0.60)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>256</td>
<td>100</td>
<td>51.6</td>
</tr>
</tbody>
</table>

* among all C4 (C4A plus C4B) allotypes

** A12, A92-Ch1;   # B94, B12, B15, B3, B6, B7

** mutant C4, no C4 protein produced from a C4A or C4B gene

*** C4Q0 mutant caused by 2-bp insertion at codon 1213 from exon 29

**Table 2.2** Phenotypes of complement components C4A and C4B in a healthy Hungarian population.
In the entire study population * \( n = 128 \) 

In subjects with bimodular / bimodular RCCX modules * \( n = 80 \) 

In subjects with at least four RCCX modules in a diploid genome * \( n = 99 \) 

<table>
<thead>
<tr>
<th></th>
<th>Unadjusted</th>
<th>Adjusted +</th>
<th>Unadjusted</th>
<th>Adjusted +</th>
<th>Unadjusted</th>
<th>Adjusted +</th>
</tr>
</thead>
<tbody>
<tr>
<td>C3</td>
<td>5.384</td>
<td>2.580</td>
<td>4.978</td>
<td>1.904</td>
<td>4.398</td>
<td>1.733</td>
</tr>
<tr>
<td></td>
<td>(&lt;0.001)</td>
<td>(0.077)</td>
<td>(0.014)</td>
<td>(0.356)</td>
<td>(0.010)</td>
<td>(0.318)</td>
</tr>
<tr>
<td>C4</td>
<td>5.249</td>
<td>5.049</td>
<td>8.812</td>
<td>8.066</td>
<td>8.801</td>
<td>8.645</td>
</tr>
<tr>
<td></td>
<td>(0.073)</td>
<td>(0.059)</td>
<td>(0.028)</td>
<td>(0.031)</td>
<td>(0.013)</td>
<td>(0.009)</td>
</tr>
</tbody>
</table>

* partial regression coefficients are shown; p-values are shown in parenthesis.

+ adjusted to age, gender, cholesterol and triglyceride levels

**Table 2.3** Linear regression analysis of the relationship between BMI and serum C3 and C4 concentrations in healthy Hungarian subjects.
CHAPTER 3

MOLECULAR BASIS OF COMPLETE COMPLEMENT COMPONENTS C4A AND C4B DEFICIENCIES IN HUMAN KIDNEY DISEASES AND SYSTEMIC LUPUS ERYTHEMATOSUS (SLE)

3.1 ABSTRACT

Complete deficiencies of human complement components C4A and C4B are strongly associated with infectious, kidney and autoimmune diseases. Among the 26 patients with complete C4A and C4B deficiencies whose clinical features were documented, the molecular basis of the genetic defects has only been elucidated in 5 subjects. To determine the molecular bases of complete complement components C4A and C4B deficiencies in patients with kidney or autoimmune diseases, four families with 7 complete C4 deficiency patients from the European Alpine region were recruited after informed consent. Complement C3 and C4 protein polymorphisms were determined by immunofixation of EDTA-plasma. C4A and C4B mutant genes were characterized by genomic Southern blot analysis and by long range mapping using pulsed field gel electrophoresis. The C4A and C4B mutant genes were amplified by long range PCR. DNA sequences for all 41 exons in each mutant gene were determined. The 7 patients are either homozygous in HLA A24 Cw7 B38 DR13 or HLA A30 B18 DR7. They have SLE, recurrent hematuria with membranous nephropathy, or Henoch-Schoenlein purpura. Molecular genetic experiments revealed a single, long mutant C4A gene in three patients.
with the HLA B38-DR13 haplotype. On the other hand, the HLA B18-DR7 haplotype has an unusual bimodular RCCX module with two short mutant C4B genes and two steroid CYP21B genes. Sequence determination revealed a novel 2-bp deletion in exon 13 in the mutant C4A genes among the three patients with the B38-DR13 haplotype. The identical G to A nucleotide substitution at the donor site of the splice junction for intron 28 was found in both of the two short C4B mutant genes from all four patients with HLA A30 B18 DR7. Definitive SSP-PCR and RFLPs have been created to facilitate screening of C4 gene mutations in SLE patients. There are two hotspots of deleterious mutations in C4AQ0 and C4BQ0: one is located at exon 13; the other is located within a 2.6 kb region spanning between exons 20 to 29. Screening of C4 mutations would help elucidate the genetic basis of autoimmune and kidney diseases.
3.2 INTRODUCTION

In the past thirty years, complete complement components C4A and C4B deficiencies have been identified and studied clinically in 13 males and 13 females from eighteen families of different racial backgrounds (Hauptmann et al 1988, Yang et al 2004). Although fifteen HLA haplotypes were present in these patients, close to three-quarters of them were homozygous in HLA alleles. All but one complete C4-deficient subjects experienced symptoms related to immune complex clearance disorders such as SLE (Lokki et al 1998), lupus-like disease (Yu et al 1986), glomerulonephritis or kidney disease (Chung et al 2002b). Many patients had recurrent microbial infections. At least four patients died at young ages because of severe infections or organ failure during lupus flares.

The human complement C4 genes are located at the class III region of the major histocompatibility complex (MHC) on the short arm of chromosome 6. The human C4 locus is remarkably complex, which is featured by (a) segmental duplication with mono-, bi-, tri- and quadri-modular RP-C4-CYP21-TNX (RCCX) in the MHC; (b) a gene size dichotomy with long and short genes; (c) two classes of polymorphic protein products for C4A and C4B (Yu et al 2003). The outcome is a sophisticated genetic diversity with multiple length variants of RCCX modules with long or short C4 genes coding for C4A and/or C4B (Chung et al 2002b, Mauff et al 1998, Yang et al 2003, Yu et al 2003). Among different individuals in a population, 2 to 7 (and possibly 8) C4 genes may be present in a diploid genome, leading to a 3-5 fold variation in plasma C4 protein concentrations and the presence of multiple allotypes for C4A and C4B proteins (Yang et al 2003). C4A is characterized by higher binding affinity to peptide antigens and
complement receptor CR1, a longer half-life of the activated molecule, a relevant role in immunoclearance and possibly a link between innate and adaptive responses. The activated C4B has a fast reaction rate towards hydroxyl group containing antigens, a short half-life and is important in propagating the complement activation pathways in a temporally and spatially specific manner. The differential chemical reactivities between C4A and C4B are attributable to the C4A/C4B isotypic residues in modulating the covalent binding activities of the thioester carbonyl group to its substrates (Dodds et al 1996, Ebanks et al 1992, van den Elsen et al 2002, Yu et al 1986). Considering the relevant roles of C4A and C4B in immunoclearance, memory, and effector functions of the humoral immune response, it is reasonable to note that a deficiency of C4A or C4B is frequently associated with infectious and/or autoimmune diseases.

Elucidation of the molecular basis of complete C4A and C4B deficiencies would help designing a comprehensive screening strategy to determine the prevalence of C4A and C4B mutations in autoimmune, infectious and kidney diseases. To date, the molecular defects of complete C4A and C4B deficiencies have been elucidated only in five subjects from three families residing in the Scandinavia and the US (Fredrikson et al 1998, Lokki et al 1998, Rupert et al 2002). These patients had single C4A mutant genes in monomodular-long (mono-L) RCCX haplotypes, or one mutant C4A and one mutant C4B genes from bimodular LS haplotypes. Among the molecular defects found, a 2-bp TC-insertion into codon 1213 has been detected in C4A mutants from different families and a C4B mutant from HLA A2 Cw7 B39 DR15 (Lokki et al 1998). In addition, a C-nucleotide deletion at codon 811 was discovered in a long C4A mutant gene of HLA A30 B18 DR3 (Fredrikson et al 1998), and another C-nucleotide deletion at codon 522 was
detected in a short C4B mutant gene from HLA A2 B12 DR6 (Rupert et al 2002). In this study, clinical histories of 7 patients with complete C4A and C4B deficiencies from four European families are described together with a detailed molecular genetic analyses to determine the RCCX modular variation and defects of the C4A and C4B mutant genes.
3.3 MATERIALS AND METHODS

3.3.1 Human subjects and peripheral blood samples

Seven complete C4 deficient patients from four unrelated families residing at the Alpine region close to border of Austria and northern Italy were recruited for this study, after informed consent. Parents from patient families 1 and 3 were also included. Peripheral blood samples were used to isolate genomic DNA and EDTA plasma, after established protocol (Yu et al 2002).

3.3.2 Allotyping of human complement components C4A and C4B proteins

EDTA plasma samples were digested with neuraminidase and carboxylpeptidase B, resolved by high voltage agarose gel electrophoresis, and fixed by goat antisera against human C4 after standard protocol (Sim and Cross 1986). The C4A and C4B electrophoretic variants were stained by Simply Blue (Invitrogen). Complement C3 protein polymorphisms in EDTA plasma were determined by the immunofixation technique similar to that for C4, using goat anti-C3 antisera.

3.3.3 Genomic RFLPs

Three RFLP strategies were employed to elucidate the number and size of C4A and C4B genes in each human subject. TaqI genomic Southern blot analysis was applied to elucidate the RP-C4-CYP21-TNX modular length variants, especially the presence of long and short C4 genes linked to RP1 or RP2, and the relative quantities of CYP21A and CYP21B, and of TNXA and TNXB. PshAl and PvuII genomic RFLP analysis was applied to determine the relative dosages of C4A and C4B genes, using a C4d probe spanning exons 22-25 for hybridization. Finally, a long-range mapping technique was
applied to give independent and confirmative information of the RCCX modules. *PmeI*
digested genomic DNA trapped in agarose plugs was resolved by pulsed field gel
electrophoresis (CHEF Mapper, Biorad) and subjected to Southern blot analysis using a
C4d specific probe (Chung et al 2002c).

### Specific oligonucleotide primers used in this study

For long range PCR of DNA fragments for cloning:

- **C4E1.5** 5’ TCC AAG AGA GGT TAG ATC CG 3’
- **C4E9.3** 5’ CTG GAG ACT AAT GAT GGC T 3’,
- **YYE10** 5’ GGA GGC AGA GC T CAC ATC CTG GTA 3’
- **Y303IN** 5’ CAG GAA GAA GTG CTG CGG GGA 3’,
- **C4E29.52** 5’ GCT CTT CTC CCT GCC TTC CT 3’,
- **C4E41.3** 5’ TGG TCC CAG GCT GTG TTC AT 3’,
- **C4E25.3** 5’ CAG GTG CTG CTG TCC CGT GA 3’,
- **Y23IN** 5’ CTC TGA CAC AGA GTC CTC AGA CC 3’.

For mutation detection using sequence-specific PCR:

- **C4E13D5** 5’ ATC CCG AGG GCA GAT CGT TC 3’
- **C4E14.3** 5’ CTG GCC CAT GTT GAG GGG CT 3’
- **C412F** 5’ CTA CTA CCC TAT TCG CAC CT 3’
- **C4E13D3** 5’ GGG CTC TCG ATT CAT GAA CGA 3’
For genetic polymorphism in two short C4B mutant genes:

E95 5’ CCC TGG AGA AGC TGA ATA TGG 3’

C4I113 5’ CCA TGG ATC CTT GGG ACC CCA 3’.

3.3.5 PCR amplification, cloning and sequencing of long C4A mutant gene from HLA A24 B38 DR13

The long C4A mutant gene was amplified in three fragments using PCR. The first fragment was 2.3 kb spanning exons 1 to 9 using primer set C4E1.5 and C4E9.3. The PCR conditions were: 1 cycle at 94°C for 2 minutes; 33 cycles at 94°C for 45 seconds, 58°C for 45 seconds, and 72°C for 3 minutes; 1 cycle at 72°C for 10 minutes. The second fragment was 6.6 kb covering exons 10-30, using primer set YYE10 and Y303IN. The PCR conditions were: 1 cycle at 98°C for 2 minutes; 40 cycles at 98°C for 45 seconds, 66°C for 60 seconds, and 72°C for 9 minutes; 1 cycle at 72°C for 10 minutes. The third fragment was 5.7 kb corresponding to exons 29 to 41, using primers C4E29.52 and C4E41.3 for amplification. The PCR conditions were: 1 cycle at 94°C for 3 minutes; 8 cycles at 94°C for 45 seconds, 64°C to 60°C for 60 seconds, and 72°C for 9 minutes; 30 cycles at 94°C for 45 seconds, 59°C for 60 seconds, and 72°C for 9 minutes; 1 cycle at 72°C for 15 minutes. All PCR reactions were performed with Failsafe PCR amplification kit (Epicentre). The 2.3 kb exon 1-9 DNA fragment was purified by PCR purification kit (Qiagen) and directly sequenced. The 6.6 kb exons 10-30 fragment and the 5.7 kb exons
29-41 fragment were purified by gel filtration, cloned into the pCR4-TOPO vector (Invitrogen) and then sequenced. Sequencing reactions were performed using Version 3 Big-Dye kit (Applied Biosystem). The sequencing products were purified by spinning device (EDGE), vacuum-dried and processed by a 16-channel capillary sequencing machine (Applied Biosystem) operated by the Sequencing Core Facility of the Columbus Children's Research Institute. DNA sequences were assembled and analyzed by GCG software. The polymorphisms and mutations discovered in the plasmid clones were further confirmed by direct sequencing of the original genomic PCR products.

3.3.6 SSP-PCR (Sequence-Specific Primer PCR) for the 2-bp deletion at exon 13

The presence of the 2-bp deletion at exon 13 of C4 gene was detected in SSP-PCR reactions using primers C4E13D5 and C4E14.3, or using primers C412F and C4E13D3. PCRs were performed using Failsafe PCR amplification kit (Epicentre). The PCR conditions were: 1 cycle at 94°C for 3 minutes; 30 cycles at 94°C for 30 seconds, 60°C for 45 seconds, and 72°C for 1 minutes; 1 cycle at 72°C for 10 minutes.

3.3.7 PCR amplification, cloning and sequencing of two short C4B mutant genes from HLA A30 B18 DR7

The short C4B mutant genes were amplified using two PCR reactions and cloned into pCR4-TOPO vector. The first fragment of 7.3 kb spanning exons 1-25 was amplified with primers C4E1.5 and C4E25.3, while the second fragment of 7.4 kb spanning exons 23-41 was amplified with primers Y23IN and C4E41.3. Both PCR reactions were performed using Failsafe PCR amplification kit (Epicentre). The PCR condition were: 1 cycle at 98°C for 2 minutes; 8 cycles at 94°C for 45 seconds, 64°C to 60°C for 60 seconds,
and 72°C for 9 minutes; 30 cycles at 94°C for 45 seconds, 59°C for 60 seconds, and 72°C for 9 minutes with a increase of 10 seconds per cycle; 1 cycle at 72°C for 15 minutes. The isolated plasmids from both PCR products were purified and sequenced to completion. To separate these two mutant C4B genes, multiple clones from each group were sequenced. The polymorphisms and mutations were further confirmed by direct sequencing of PCR products.

### 3.3.8 SSP-PCR-MboI RFLP to detect mutation at the 5’ splice site of intron 28

SSP-PCR was used to determine the g→a mutation in intron 28 of C4B gene using primers I27F and MBO-28R. PCR reaction was performed using Failsafe PCR amplification kit (Epicentre). The PCR condition was: 1 cycle at 96°C for 3 minutes; 33 cycles at 96°C for 45 seconds, 62°C for 45 seconds, and 72°C for 2 minutes; 1 cycle at 72°C for 5 minutes. The PCR products were digested with restriction enzyme MboI at 37°C and resolved by electrophoresis with a 1.5% to 2% agarose gel.

### 3.3.9 PCR for exon 9-11 to segregate two mutant C4B genes

To detect the T/C polymorphism in intron 9, exon 9-11 in the patients with two mutant C4B genes was amplified using primers E95 and C4I113. The PCR conditions were: 1 cycle at 94°C for 2 minutes; 35 cycles at 94°C for 45 seconds, 58°C for 60 seconds, and 72°C for 2 minutes; 1 cycle at 72°C for 7 minutes. The PCR products were digested by HinPI at 37°C, and resolved by agarose gel electrophoresis.
3.3.10 Additional declarations

We are indebted to the patients and family members who participated in this study. The peripheral blood samples were withdrawn locally and sent to us by our collaborators Dr. Karl Lhotta, who also contributed to the clinical features of the patients. PmeI PFGE of all the available family members were performed by another graduate student Erwin Chung. The C4 and C3 immunofixation experiments were determined by our research associate Bi Zhou. The –308A/G polymorphisms were determined by Yee ling Wu. All the other experiments were done by the author.

A manuscript of this work has been submitted for publication. The citation is as follows: Yang Y, Lhotta K, Chung EK, Eder P, Neumair F, Yu CY: Molecular basis of complete complement components C4A and C4B deficiencies in human kidney diseases and systemic lupus erythematosus (SLE).
3.4 RESULTS

3.4.1 Patients’ clinical histories

Seven complete C4 deficiency individuals from four independent families resided at Sudtirol, an Alpine area in northern Italy close to Austria, were recruited for the current study. The clinical histories of these patients are listed on Table 3.1 and described as follows.

Family 1

1P: The 42-year old male patient suffered from severe Henoch Schoenlein purpura at the age of seventeen with involvement of the skin, intestine and kidneys (Lhotta et al 1990, Lhotta et al 1993a, Tappeiner et al 1978). Six years later he developed macrohematuria and the nephrotic syndrome. A renal biopsy showed mesangial glomerulonephritis with fibrous crescents and tubular atrophy. At the age of 23, hemodialysis had to be started. One year later he received a renal allograft. After two years hematuria and proteinuria were noted and a biopsy of the transplanted kidney showed recurrence of mesangial glomerulonephritis and chronic allograft nephropathy. Five years after transplantation dialysis had to be restarted. After eight years of hemodialysis he received a second renal allograft. Six years later his serum creatinine is 130 µmol/l and urinalysis is normal without signs of recurrent disease. Recent treatment regimen includes tacrolimus, azathioprine and prednisolone.

The 37-year-old brother of the patient also has complete C4 deficiency. He is unavailable for clinical investigation but is reported to be healthy.
Family 2

2P: The now twenty-year old boy presented at the age of ten with recurrent attacks of fever, vomiting and macrohematuria. A renal biopsy showed mild mesangioproliferative glomerulonephritis with immune deposits in the mesangium. He was treated with low-dose steroids and amoxycillin (Lhotta et al 1996a, Lhotta et al 2001). At the age of fifteen, after a wound infection, he developed a nephrotic syndrome with proteinuria of ten gram per day. Renal histology revealed a membranous-type glomerulonephritis with huge epimembranous immune deposits. He responded well to treatment with intravenous immunoglobulin 1 gram per kilogram bodyweight monthly for ten months with reduction of protein excretion below one gram. However, after that treatment was stopped, proteinuria recurred and remained unresponsive to immunoglobulin infusion. Treatment with mycophenolate mofetil was initiated and a partial response with reduction of proteinuria to 2.5 gram per day was achieved. Renal function remains normal (Lhotta K., manuscript submitted for publication).

Family 3

The three children of the family suffered primarily from proliferative glomerulonephritis. Two of them developed end-stage renal failure. The third sibling had life-threatening cerebral involvement.

3P1: This now 33-year old female patient developed a lupus-like disease at age six. She had erythema of the face, hands and arms. Urinalysis showed microhematuria and proteinuria of three gram per day and hypertension. A renal biopsy showed a membranoproliferative-like glomerulonephritis with immune complex deposition predominantly in the mesangium. She received azathioprine and steroids. Despite
treatment she developed slowly progressive chronic renal failure. Hemodialysis had to be started when she was 26 years of age. After five years on hemodialysis she received a renal allograft. Two years later she has normal transplant function and no signs of recurrence of glomerulonephritis in the allograft. Her current immunosuppressive regimen consists of cyclosporin A, mycophenolate mofetil and prednisolone (Lhotta et al 1993b).

3P2: This male patient is now 26 years old. At the age of five he developed lupus-like skin lesions, microhematuria and proteinuria of two gram per day. When he was nine years old a renal biopsy was performed, which showed severe membranoproliferative-like glomerulonephritis with mesangial proliferation. All glomerular capillaries showed proliferative changes of variable severity. Despite immunosuppressive treatment with azathioprine and prednisolone renal function deteriorated. Cyclophosphamide bolus therapy stabilized renal disease for some time, but at the age of sixteen the patient was started on hemodialysis. After two years a cadaveric renal transplantation was performed. He was treated with tacrolimus, azathioprine and steroids. Because of proteinuria and increasing serum creatinine levels five years after transplantation a graft biopsy was performed. The biopsy showed no signs of recurrence of lupus nephritis but chronic allograft nephropathy. Six years after transplantation hemodialysis was again necessary. After three months on hemodialysis he suffered from Aspergillus fumigatus meningitis. The patient was treated with liposomal amphotericine B and made a full recovery. Currently patient 3P2 is on hemodialysis and on the waiting list for a second renal transplantation (Lhotta et al 1993b, Moling et al 2002, Tappeiner et al 1982).
**3P3**: This female patient is now 23 years of age. At five years old she was noted to have hematuria and proteinuria of four grams per day. A renal biopsy showed mesangial and focal endocapillary proliferative glomerulonephritis and the diagnosis of systemic lupus erythematosus (SLE) was made. Treatment with azathioprine and prednisolone was begun and resulted in marked improvement of proteinuria. She was well until the age of 22 when she developed a febrile illness with a facial maculopapular rash. A skin biopsy revealed vasculitis with immune complex deposits. Her mental status deteriorated rapidly and MR imaging showed severe cerebral vasculitis. The patient’s condition was unresponsive to high-dose steroids, plasma infusion and exchange and intravenous immunoglobulins. However, she responded to immunoadsorption treatment and mycophenolate mofetil and an almost complete recovery of cerebral function was achieved. At present she is maintained on mycophenolate mofetil and low-dose steroids without signs of glomerulonephritis or vasculitis (manuscript submitted).

Family 4

The prominent clinical presentation in the patients of that family is lupus-like skin disease. Patients also suffered from mild glomerular disease.

**4P1**: The male patient is now twenty-nine years old. At the age of five he suffered from acute oliguric renal failure. A renal biopsy revealed mild mesangial glomerulonephritis. Steroid treatment led to complete resolution and normalization of renal function. Serum creatinine and urine analysis were normal at the time of this report. The patient developed skin manifestations of systemic lupus erythematosus primarily on the face. For that condition he was treated with hydroxychloroquine 200 mg daily and low-dose prednisolone (Lhotta et al 1993b, Tappeiner et al 1982).
4P2: The female patient now is forty years of age. When she was two years old she developed recurrent attacks of fever, a rash on the face, trunk and extremities and oral ulcers. She was treated with systemic and later, topical steroids. Skin lesions became atrophic and scarring. At the age of 24 a renal biopsy was obtained because of microscopic hematuria. Histologic studies showed a mesangial glomerulonephritis with increase in mesangial matrix and extensive mesangial immune complex deposition. At present renal function is normal and there is no hematuria or proteinuria. The patient continued to suffer from skin involvement. She required plastic surgery with autologous skin transplantation to her chin at age 29. Her current medication is azathioprine 100 mg, hydroxychloroquine 200 mg and low-dose steroids in order to control skin disease (Lhotta et al 1993b, Tappeiner et al 1982).

A brother of these two patients died at the age of three from cerebral vasculitis and sepsis caused by Staphylococci and Streptococci (Tappeiner et al 1982).

3.4.2 RCCX modules, C4A and C4B mutant genes in the seven complete C4 deficiency patients

HLA typing revealed two common haplotypes were present among the seven patients described above. Patients 2P, 4P1 and 4P2 were homozygous with HLA A30 B18 DR7; patients 1P, 3P1, 3P2 and 3P3 were homozygous with HLA A24 B38 DR13 (Lhotta et al 1990, Lhotta et al 1993b, Lhotta et al 1996b). Immunofixation experiments of EDTA-plasma confirmed the complete absence of complement C4A and C4B proteins in these seven patients (panel A, Figure 3.1). The results also showed that the mother of 1P expressed C4A3, A2, and both parents of patients 3P1, 3P2 and 3P3 expressed C4A3, C4B1. Complement C3 proteins were detectable in all patient samples and their relatives...
using the same EDTA-plasma samples and immunofixation technique (panel B, Figure 3.1), suggesting the absence of C4 proteins were not likely caused by protein degradation.

The RCCX structures of the patients were determined by TaqI RFLP (panel C, Figure 3.1) and further confirmed by PmeI-PFGE (panel D, Figure 3.1) of genomic DNA samples. The dosages of C4A and C4B genes were determined by PshAI-PvuII RFLP (panel E, Figure 3.1).

The three patients with homozygous HLA A24 B38 DR13, 2P, 4P1 and 4P2, had the identical TaqI restriction patterns that are characteristic of the monomodular-long (mono-L) RCCX structures. As shown in left panel C of Figure 3.1, each patient had RP1 linked to a long C4 gene (7.0 kb), followed by a CYP21B gene (3.7 kb) and a TNXB gene (2.5 kb) in the RCCX module. PmeI PFGE revealed the presence of a 113 kb fragment, confirming the presence of homozygous (L/L) RCCX structure (left panel D, Figure 3.1). PshAI-PvuII RFLP using a C4d probe further revealed that the mutant C4 gene present in 2P, 4P1 and 4P2 are all C4A, as only the 1.7 kb C4A-specific restriction fragments were detectable. In essence, there is a single, long C4A mutant gene (C4AQ0) present in the monomodular RCCX structure with RP1-C4AQ0 (L)-CYP21B-TNXB between HLA-A24, -B38 and HLA-DR13.

For the four patients with homozygous HLA A30 B18 DR7 haplotypes from Families 1 and 3, TaqI RFLP showed the presence of bimodular RCCX structures. These structures were characterized by RP1 linked to a short C4 gene (6.4 kb), followed by CYP21B (3.7 kb) and TNXA (2.4 kb) in the first module, and then RP2 linked to another short C4 gene (5.4 kb), followed by CYP21B (3.7 kb) and TNXB (2.5 kb) in the second
module. The presence of such bimodular RCCX structure with two short C4 genes were further confirmed by the 139 kb *Pmel* fragment in the PFGE (panel D, Figure 3.1).

*PshAl-PvuII* RFLP of the C4d region revealed that the C4 genes in 1P, 3P1, 3P2 and 3P3 were all C4B, as only the 2.2 kb C4B-specific restriction fragments were detectable. Therefore, the four patients with homozygous HLA A24 B38 DR13 haplotypes contained two short C4B mutant genes (C4BQ0) and no C4A genes.

*TaqI* RFLPs showed that CYP21A was present in none of the seven complete C4 deficiency patients. This is unusual as a bimodular RCCX structure often contains a CYP21A together with TNXA-RP2 between the two C4 genes. As shown in Figure 3.1, no CYP21A specific *TaqI* fragment is detectable in all of the complement C4A and C4B deficient patients engaged in this study. The relative band intensities of CYP21B:CYP21A in the mother of family 1 and the parents of family 3 both showed a ratio of 3:1, which suggested a configuration of CYP21B-CYP21B in the A30 B18 DR7 haplotype. To further confirm the absence of CYP21A, *BsaI* RFLP was performed. The 8-bp deletion in exon 3 of CYP21A creates a new *BsaI* restriction site. As shown in Figure 3.2, no CYP21A specific *BsaI* fragments were detectable in the complete C4-deficiency patients. Based on these genotypic and phenotypic analyses, the RCCX modules in HLA A30 B18 DR7 is interpreted as bimodular SS with the following configuration: RP1-C4BQ0 (S)-CYP21B-TNXA-RP2-C4BQ0 (S)-CYP21B-TNXB.

Genotype analysis of the mother from patient 1P revealed LS / SS heterozygous RCCX structures (1M, panels A and B, Figure 3.1). In addition to the SS-structure with two mutant C4B genes, her other two C4 genes from the LS-structure coded for C4A3 and C4B5 proteins (panel D, Figure 3.1). The parents of patients 3P1, 3P2 and 3P3 were
also heterozygous in RCCX structures. The mother of Family 3 (3M) had LL/SS haplotypes and the father 3F had LS/SS haplotypes (panels A and B, Figure 3.1). Both the LL haplotype from 3M and the LS haplotype from 3F expressed C4A3 and C4B1.

3.4.2.1 A 2-bp deletion in exon 13 of the long C4A mutant gene from HLA A24 Cw7 B38 DR13

Sequencing determination and analyses was first initiated on patient 2P at the polymorphic C4d region that spanned 2.3 kb. It was found that the mutant gene contained sequences characteristic of C4A, which included the sequences coding for D1054, PCPVLD 1101-6, and VDLL 1188-91. In addition, it had the typical indels in intron 28 and 29 that are present in long C4A genes. However, no deleterious mutations were detected at the C4d genomic region of the mutant gene.

To identify the nucleotide mutations contributing to the non-expression of the complement protein, the long C4A mutant gene from patient 2P was amplified by PCR in three fragments. These fragments corresponded to exons 1-9 (2.3 kb), exons 10-31 (6.6 kb) and exons 29-41 (5.7 kb). The 6.4 kb endogenous retrovirus HERV-K(C4) located in intron 9 was not included. The three amplified fragments were cloned and sequenced using 60 primers that allowed sequence determination of all 41 exons and their intervening introns in both orientations (Rowen et al 1997a, Yu 1991, Yu et al 2003).

When compared with the annotated sequence for C4A and C4B (Yu 1991, Yu et al 2003), ten nucleotide substitutions or indels present in the C4A mutant gene that deserved special attention are listed in Table 3.2. The most conspicuous change is a deleterious 2-bp deletion at the sequence for codon 497 from exon 13 (panel B, Figure 3.3). At nucleotide 9968-9 (or 3594-5 for a short C4 gene), the sequence from patient 2P was
missing the GT dinucleotide. Such deletion would change the protein reading frame and generate terminations at codons 607 and 613 from exon 15 (panel C, Figure 3.3). Two other novel nucleotide changes were C9775T that was a synonymous mutation at the sequence for codon 476 from exon 12, and the c14701t substitution at intron 28.

Subsequently, genomic fragments at exon 13 from patients 2P, 4P1 and 4P2 were independently amplified by PCR and subjected to sequencing. The identical mutation was found in genomic DNA samples from all three patients with HLA A24 Cw7 B38 DR13.

3.4.2.2 Screening of this 2-bp deletion in exon 13

To facilitate screening of the 2-bp deletion at exon 13 from genomic DNA samples, specific forward (E13D5) and reverse (E13D3) PCR primers were designed for two independent experiments. The application of E13D5 and E14 yielded a 490 bp fragment from patients 2P, 4P1 and 4P2. The application of 12f and E13D3 yielded 450 bp fragments from the same subjects. In each set of experiment, the positive control was a 757-bp fragment amplified by 21A5 and 21A3 (panel A, Figure 3.5).

3.4.2.3 Identical mutations in the short C4B mutant genes from HLA A30 B18 DR7

Genomic DNA fragments corresponding to the two short C4B genes from patient 1P were amplified together by PCR in two independent experiments as shown in Figure 3.4B. Each of these DNA fragments was sequenced to completion. Variant sequences were identified by comparing with C4B sequences in public databases and listed on Table 3.3.
Nine novel nucleotide changes were detected in the mutant C4B genes. However, none of these changes were located in the coding sequences. Peculiarly, five novel single nucleotide mutations clustered in intron 19. The remaining four mutations were present in introns 20, 28, 30 and 31, respectively. Remarkably, the g→a substitution at intron 28 was present at the intron donor site (position 8127; panel D, Figure 3.4). Such substitution (i.e., gt→at) would abrogate the correct splicing of C4 RNA transcripts. A new potential splice junction is present 7 nucleotides downstream of original donor site. If the C4 heteronuclear RNA were spliced according to this cryptic donor site, a new termination codon TAA would be generated 9-nucleotides downstream of Gly-1206 (panel D, Figure 3.4).

The homogeneity of the genomic DNA sequences at the 3’ region of exon 28 from patient 1P (right panel C, Figure 3.4) suggested that both short C4B genes in the PCR product had the same mutation. In contrast, direct sequencing of the corresponding genomic region amplified for the maternal DNA sample of subject 1M (the mother of patient 1P) yielded both “g” and “a” sequences. The later was expected because subject 1M had two functional C4 genes coding for C4A3 and C4B5, in addition to the two mutant C4B genes (Figure 3.1).

Sequencing of the mutant C4B genes from patients 3P1, 3P2 and 3P3 also revealed homogeneous and identical sequences with the g→a substitution at nucleotide 8127, the 5’ splice junction or the donor site of intron 28.

3.4.2.4 Screening of mutation g8127a

To facilitate screenings of genomic DNA samples for the g→a mutation at the donor site of intron 28 (nucleotide 8127), a new PCR strategy was created that employed
a reverse primer with one mutagenized nucleotide to create a novel MboI restriction after amplification across the intron 28 donor site. As shown in Figure 3.5, the DNA fragments corresponding to intron 27-intron 28 from 1P, 3P1, 3P2 and 3P3 were homogeneous with the presence of g8127a mutation. As expected, the sample from subject 1M was heterozygous and the normal control gave rise to the uncleaved fragment only.

3.4.2.5 Two different short C4B genes in HLA A30 B18 DR7

Sequence determination of the amplified DNA fragments showed that both short C4B genes had virtually identical sequences except one single nucleotide at intron 9. At position 2601, both “t” and “c” nucleotides were detectable, suggesting a possible diversion from the two mutant C4B genes. The C4BQ0 gene with 2601c is detectable by restriction enzyme HinPI, which recognizes DNA sequence “gcgc”. To confirm the presence of two different C4BQ0 genes in the HLA A30 B18 DR7 haplotype, a 931-bp genomic DNA fragments spanning exon 9 to exon 11 was amplified by PCR and digested by HinPI. If an HinPI site is present, the restriction digested products would be 585 bp + 346 bp in size. Figure 3.5 showed the result of such an experiment. With respect to the presence of 2601c, the control sample (lane C, panel C, Figure 3.5) was homogeneously positive. On the other hand, patient members of the family 1 and family 3 all yielded heterogeneous results as both 2601c-positive and 2601c-negative fragments were detected.
3.5 DISCUSSION

Here we described the clinical histories and the molecular bases of seven patients with complete complement C4 deficiencies. These patients came from four independent families and they are homozygous in two HLA haplotypes. The first group with HLA A24 B38 DR13 contains the monomodular RCCX haplotype with a single long C4A mutant gene. In other words, no C4B gene is present. The molecular defect leading to the absence of C4A protein production is a 2-bp deletion at exon 13. The three patients with this molecular defect have severe skin lesions, SLE and frequent infections. The second group with HLA A30 B18 DR7 contains the bimodular RCCX with two short mutant C4B genes. No C4A gene is present in this haplotype. The molecular defects are probably a point mutation at the donor site splice junction of intron 28 in each of the two C4B mutants. The four patients with this defect have severe kidney disorders and SLE. Three of these patients had end stage kidney failures and required hemodialysis and kidney transplantations. The graft kidney function in one patient (1P) deteriorated at year 2 and required hemodialysis again at year 5 post-transplantation. A second renal transplantation was eventually performed. The SLE disease in another patient (patient 3P3) progressed to a life-threatening cerebral vasculitis, an illness that claimed the life of another complete C4 deficiency patient with whom we studied earlier (Rupert et al 2002). A complex treatment regimen, which included the application of mycophenolate mofetil to suppress B cell proliferation and a potential alloimmune response, and the infusion of fresh frozen plasma that contained complement C4 but depleted with immune complexes through immunoadsorption, appeared to have reversed the disease course and the patient regained much of her CNS function (Lhotta K., 2004)
The roles of complement C4A and C4B in immunity, autoimmunity and kidney physiology are intriguing. The current thoughts are that complete absence of C4 proteins probably impaired the clearance of immune complexes and apoptotic materials, which contributes to inflammatory and vasculitic lesions in various organs including the skin and the kidneys. It is postulated that the presence of complement C4-decorated self antigens would facilitate the deletion of autoreactive B cells in the bone marrow involved in the central tolerance, while the deposition of activated C4 on foreign antigens facilitates the activation of antigen specific B cells and enhances the class switching of immunoglobulins in the peripheral lymphoid system (Carroll 1998a, Ochs et al 1983). Thus, it is reasonable for a connection between complete C4 deficiency with systemic autoimmune diseases and kidney disorders. On the other hand, multiple investigators observed a drop of complement C4 protein levels (due to complement consumption) in patients with lupus nephritis (Buyon et al 1992, Ho et al 2001, Manzi et al 1996). Also, the deposition of C4d, which is a split product of inactivated C4 containing the thioester residues, is found to be one of the most consistent markers for acute and chronic renal graft rejections caused by humoral alloimmune response (Feucht et al 1993). These phenomena probably reflect the tissue injuries caused by effector functions of C4 in the complement pathways.

To date, six deleterious mutations in C4A or C4B genes have been detected in 12 human subjects with complete C4A and C4B deficiencies. All except one of those mutations are 1-bp or 2-bp insertions or deletions (indels) in coding sequences that lead to frameshift and non-sense mutations (Figure 3.6). In the short C4B genes of HLA A30 B18 DR7, a G→A transition at the donor site of intron 28 is found in both mutant C4...
genes of the bimodular RCCC haplotype. Such change would lead to aberrant splicing of the C4 heteronuclear RNA and therefore introduce changes in reading frames and stop codons. SSP-PCR and SSP-PCR plus RFLP techniques to screen mutations of C4 genes have now been created and this would help clarify the role of C4A and/or C4B deficiencies in infectious and autoimmune disease patients. However, it is of interest to point out that deleterious nonsense mutations tend to be race or ethnic group specific (Burchard et al 2003). For example, the presence of the 2-bp insertion in exon 29 of the C4A genes have been detected in the healthy Caucasians (Blanchong et al 2000) and in White and Black SLE patients (Sullivan et al 1999). However, such mutation is not detectable in east Orientals (Man et al 2003). While considerable progress has been made in understanding the molecular basis of C4A and/or C4B deficiencies in European and Northern American Caucasians, very little or no knowledge is available on the basis of C4 deficiencies in any other ethnic groups.

As established previously, a bimodular RCCX is regularly characterized by the presence of an RP1 gene, followed by C4A, the pseudogene CYP21A, gene fragments TNXA and RP2, C4B, the steroid 21-hydroxylase CYP21B, and then extracellular matrix protein TNXB. Two major bimodular RCCX haplotypes are present in the Caucasians, LL and LS. With some exceptions, the first long gene usually code for C4A. The second gene may be long or short, which mostly codes for C4B and sometimes for C4A. The C4 genes and RCCX constituents in the HLA A30 B18 DR7 haplotype contain a couple of distinct features. The first is the presence of bimodular SS with two short C4B genes in a row. These two C4B genes share the identical mutations and only one nucleotide change is detectable between them. The second is the presence of CYP21B-CYP21B
instead of CYP21A-CYP21B. These phenomena could be explained by a recent gene
duplication event, or by a genetic process such as a gene conversion that homogenized
the C4 and CYP21 sequences. The linkage of the TNF2 allele with the bimodular SS in
the MHC class III region of HLA A30 B18 DR7 is also of interest. This is because the
presence of a short C4B gene linked to RP1 with –308A allele of TNFA (the TNF2
allele) is a characteristic of HLA A1 B8 DR3 in Europeans, which is described as
other bimodular SS structures including the C4A/C4B genes, CYP21A/CYP21B genes
and HLA class I and class II alleles might shed light on the origin of this unusual
haplotype.
Figure 3.1 Phenotypic (panel A) and genotypic analyses (panels C, D and E) of complement C4 in seven complete C4 deficiency patients and available family members. Complement C3 was determined by C3 immunofixation (panel B) and used as a control to show the quality of the plasma samples.
Figure 3.2 Determination of CYP21A and CYP21B genes in complete C4 deficient subjects and family members from families 1, 3 and 4 using BsaI RFLP.
A. PCR strategy to amplify the mutant C4A gene in monomodular-long RCCX

B. The 2-bp deletion in exon 13 of the C4A mutant gene

C. The 2-bp deletion in the long C4AQ0 gene causes frameshift and nonsense mutations

Figure 3.3 Monomodular (mono-L) RCCX structure, PCR amplifications and sequence determination of the mutation in long C4A gene from HLA A24 Cw7 B38 DR13. Arrows in panel B indicated the 2-bp sequence, which was deleted in the C4A mutant gene. Asterisks in panel C indicated stop codons.
Figure 3.4 Bimodular SS RCCX structures, PCR amplification and sequence determination of the mutations in the two short C4B genes from HLA A30 B18 DR7.
A. SSP-PCR to show the 2-bp deletion at exon 13 in the C4AQ0 gene of the B38-DR13 haplotype

![Image of SSP-PCR to show the 2-bp deletion at exon 13 in the C4AQ0 gene of the B38-DR13 haplotype]

B. SSP-PCR and MboI RFLP to show donor splice site mutation of intron 28 in C4BQ0 genes of the B30-DR7 haplotype

![Image of SSP-PCR and MboI RFLP to show donor splice site mutation of intron 28 in C4BQ0 genes of the B30-DR7 haplotype]

C. HinPI RFLP to show a SNP in the intron 9 of the two C4BQ0 genes

![Image of HinPI RFLP to show a SNP in the intron 9 of the two C4BQ0 genes]

Figure 3.5 SSP-PCRs to determine the novel mutations in the long C4A gene (panel A) and the two short C4B genes (panel B), and the sequence polymorphisms between these two short C4B genes (panel C).
Figure 3.6 The hotspots of deleterious mutations in C4AQ0 and C4BQ0 in humans. Five different mutations were identified among fifteen complete C4 deficient subjects.
<table>
<thead>
<tr>
<th>Patient no.</th>
<th>Sex and Age (disease onset / current)</th>
<th>Disease and Organ Involvement</th>
<th>Medical Procedures</th>
<th>Recent Medication</th>
</tr>
</thead>
<tbody>
<tr>
<td>2P</td>
<td>M; 10 / 20</td>
<td>kidney (mesangial GN; membranous GN)</td>
<td></td>
<td>mycophenolate mofetil 1.5 g prednisolone 5 mg</td>
</tr>
<tr>
<td>4P1</td>
<td>M; 5 / 29</td>
<td>SLE skin kidney (mesangial GN)</td>
<td></td>
<td>hydroxychloroquine 200 mg prednisolone 7 mg</td>
</tr>
<tr>
<td>4P2</td>
<td>F; 2 / 40</td>
<td>SLE skin skin kidney (mesangial GN) autotransplantation (chin)</td>
<td></td>
<td>hydroxychloroquine 200 mg azathioprine 100 mg prednisolone 8 mg</td>
</tr>
<tr>
<td>1P</td>
<td>M; 17 / 42</td>
<td>Schoenlein Henoch purpura kidney (mesangial GN), skin, gut</td>
<td></td>
<td>tacrolimus 6 mg azathioprine 50 mg prednisolone 5 mg</td>
</tr>
<tr>
<td>3P1</td>
<td>F; 6 / 33</td>
<td>SLE kidney (MPGN) transplantation</td>
<td></td>
<td>cyclosporine A 150 mg mycophenolate mofetil 2 g</td>
</tr>
<tr>
<td>3P2</td>
<td>M; 5 / 26</td>
<td>SLE kidney (MPGN) transplantation</td>
<td></td>
<td>hemodialysis sevelamer 4.8 g furosemide 500 mg</td>
</tr>
<tr>
<td>3P3</td>
<td>F; 5 / 23</td>
<td>SLE kidney (MPGN) skin cerebral vasculitis</td>
<td></td>
<td>mycophenolate mofetil 1 g prednisolone 4 mg</td>
</tr>
</tbody>
</table>

GN: glomerulonephritis; MPGN, membranoproliferative GN

Table 3.1 A Summary of Clinical Histories in 7 Complement C4 Deficiency Patients.
<table>
<thead>
<tr>
<th>Exon / Intron</th>
<th>Position</th>
<th>Nucleotide change</th>
<th>amino acid change (if any)</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exon 9</td>
<td>2296</td>
<td>C → A (TCT to TAT)</td>
<td>S 328 Y</td>
<td></td>
</tr>
<tr>
<td>Exon 12</td>
<td>9775 (3401)</td>
<td>C → T (GCC to GCT)</td>
<td>476 A</td>
<td>nc</td>
</tr>
<tr>
<td>Exon 13</td>
<td>9968 (3594)</td>
<td>– GT</td>
<td>nc; frameshift and premature stop codon</td>
<td></td>
</tr>
<tr>
<td>Exon 20</td>
<td>12169 (5796)</td>
<td>T → C (GTT to GTC)</td>
<td>806 V</td>
<td></td>
</tr>
<tr>
<td>Intron 20</td>
<td>12299 (5923)</td>
<td>g → a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exon 21</td>
<td>12526 (6150)</td>
<td>A→G (ACC to GCC)</td>
<td>T 888 A</td>
<td></td>
</tr>
<tr>
<td>Intron 28</td>
<td>14510 (8140)</td>
<td>g → c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intron 28</td>
<td>14514 (8144)</td>
<td>+ c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intron 28</td>
<td>14701 (8332)</td>
<td>c → t</td>
<td>nc</td>
<td></td>
</tr>
<tr>
<td>Intron 29</td>
<td>14980 (8611)</td>
<td>c → g</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.2 The sequence changes in the mutant long C4A gene in Families 2 and 4.
<table>
<thead>
<tr>
<th>Exon/Intron</th>
<th>Nucleotide number*</th>
<th>Nucleotide change (c.f. C4B short)</th>
<th>Amino acid substitution (if any)</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exon 9</td>
<td>2296</td>
<td>C $\rightarrow$ A (TCT to TAT)</td>
<td>S 328 Y</td>
<td></td>
</tr>
<tr>
<td>Intron 9</td>
<td>2601</td>
<td><strong>c or t</strong></td>
<td></td>
<td>segregates 2 C4B mutant genes</td>
</tr>
<tr>
<td>Exon 16</td>
<td>4659</td>
<td>C $\rightarrow$ T (AAC to AAT)</td>
<td>N 661</td>
<td></td>
</tr>
<tr>
<td>Intron 19</td>
<td>5546</td>
<td>t $\rightarrow$ c</td>
<td>nc</td>
<td></td>
</tr>
<tr>
<td>Intron 19</td>
<td>5653</td>
<td>a $\rightarrow$ g</td>
<td>nc</td>
<td></td>
</tr>
<tr>
<td>Intron 19</td>
<td>5666</td>
<td>g $\rightarrow$ a</td>
<td>nc</td>
<td></td>
</tr>
<tr>
<td>Intron 19</td>
<td>5753</td>
<td>a $\rightarrow$ g</td>
<td>nc</td>
<td></td>
</tr>
<tr>
<td>Intron 19</td>
<td>5761</td>
<td>t $\rightarrow$ c</td>
<td>nc</td>
<td></td>
</tr>
<tr>
<td>Exon 20</td>
<td>5793</td>
<td>T $\rightarrow$ C (GTT to GTC)</td>
<td>V 806</td>
<td></td>
</tr>
<tr>
<td>Intron 20</td>
<td>5985</td>
<td>g $\rightarrow$ a</td>
<td>nc</td>
<td></td>
</tr>
<tr>
<td>Exon 21</td>
<td>6150</td>
<td>A $\rightarrow$ G (ACC to GCC)</td>
<td>T 888 A</td>
<td></td>
</tr>
<tr>
<td>Intron 21</td>
<td>6307</td>
<td>g $\rightarrow$ a</td>
<td>nc</td>
<td></td>
</tr>
<tr>
<td>Intron 21</td>
<td>6331</td>
<td>a $\rightarrow$ g</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intron 21</td>
<td>6419</td>
<td>t $\rightarrow$ c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exon 26</td>
<td>7535</td>
<td>C $\rightarrow$ A (GGC to GGA)</td>
<td>G 1076</td>
<td></td>
</tr>
<tr>
<td>Intron 28</td>
<td>8127</td>
<td>g $\rightarrow$ a</td>
<td>nc; donor splice site</td>
<td></td>
</tr>
<tr>
<td>Intron 30</td>
<td>8920</td>
<td>c $\rightarrow$ a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intron 31</td>
<td>9576</td>
<td>g $\rightarrow$ a</td>
<td>nc</td>
<td></td>
</tr>
<tr>
<td>Intron 31</td>
<td>10248</td>
<td>t $\rightarrow$ c</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* nucleotide number is based on sequence of a short C4B gene; nc, novel change in sequence.

Table 3.3 Nucleotide Changes in the short C4B genes from Patient 1P (HLA A30 B18 DR7; RCCX - SS).
CHAPTER 4

THE ASSOCIATION STUDY OF COMPLEMENT C4, C2 AND TNFA WITH SYSTEMIC LUPUS ERYTHEMATOSUS (SLE)

4.1 ABSTRACT

Systemic lupus erythematosus (SLE) is an autoimmune disease caused by the interplay of genetic and environmental factors. The major pathogenic pathway in SLE is the formation of various autoantibodies leading to high circulating levels of immune complexes that probably initiate the various lesions in SLE. Several disease susceptibility genes located in MHC region, including early complement components C4 and C2, and TNFA, were indicated by many studies to be associated with SLE development. The difficulty of studying the genes in MHC region is caused by the strong linkage disequilibrium of some specific alleles with some common HLA haplotypes. Complement C4 is one of the most polymorphic molecules in the human genome. The complexity of C4 genetics, including gene number, gene size and protein polymorphism, makes it hard to accurately determine its variation in human SLE. In this study, the variations in total C4, C4A and C4B gene dosages and RCCX structures, plasma C4, C4A and C4B protein levels and protein polymorphisms and C4 split product C4a levels were determined using well-designed techniques in our laboratory. It was shown that the heterozygous deficiencies or low gene dosages of total C4 ($p=0.000018$) and C4A
(p=0.00041) genes were increased significantly in female Caucasian SLE patients (N=209). In addition, mono-S increased to 19.1% (in haplotypes). Although not statistically significant, the deficiencies of C4 and C4A genes were increased in African American SLE (N=56). In Caucasian SLE, the common C4A mutation due to a 2-bp insertion in exon 29 increased from 2% in controls to 5.79% in SLE. Intriguingly, protein levels of total C4, C4A and C4B decreased in SLE patients, even within the same gene dosage group. The C4 consumption, indicated by C4 split product C4a, also increased in SLE patients. Therefore, the low protein levels of total C4, C4A and C4B in SLE were caused by deficiencies of total C4 and C4A genes, C4A mutations, and low C4 expression and/or high C4 consumption. The consequence of low plasma C4 protein levels would cause inefficiency to clear immune complexes and apoptotic bodies in the circulation that may finally lead to their deposition in local tissues and subsequent tissue damages. Although C2 deficiency and TNFA –308 polymorphisms were suggested to be associated with SLE in some previous studies, our results did not support the notion. C2 deficiency was rare (only 1.58% in our SLE population), making it unlikely to be a primary and common risk factor involved in human SLE development. TNFA –308A allele was indeed increased in SLE population. However, detailed analysis indicated the increased frequency was mainly derived from the association with mono-S, suggesting TNFA was a secondary factor in lupus development. In summary, our current study revealed a pivotal role of complement C4 and C4A deficiencies in SLE development.
4.2 INTRODUCTION

SLE (Systemic Lupus Erythematosus) is a prototype of systemic autoimmune disease. The cause of the disease is multifactorial, including both genetic and environmental factors. It has the potential to involve multiple organs and systems, and its clinical manifestations are diverse and variable. The most common phenomenon of SLE is the presence of various autoantibodies that lead to high circulating levels of immune complexes (ICs). These circulating ICs probably initiate the various lesions in SLE. The deposition of ICs in tissues could lead to self-damaging complement activation and other inflammatory responses (Lahita 1999, Vyse and Kotzin 1998, Wallace and Hahn 1993).

The susceptibility to many autoimmune diseases including SLE is associated with the presence of particular haplotypes of genes in major histocompatibility complex (MHC, also termed as HLA in humans). Genes in the MHC code for three major groups of proteins, referred to as class I, II and III. There are at least 128 expressed genes in human MHC (Yu et al 2000). Human complement C4 and C2 are located in MHC class III region. It is well known that deficiencies in early components of the classical complement activation pathway, C4, C2 and C1q, C1r, C1s, are important risk factors of SLE (Carroll 1998a, Carroll 1998b, Navratil et al 1999, Sullivan 1998). In addition to C4 and C2, two other loci in the MHC region are particularly relevant. One is the HLA class II gene DR, the other is TNFA gene.

Human complement C4 is an essential component of the classical and the lectin complement activation pathways. It is one of the most polymorphic molecules in the human genome. C4 can be classified into C4A and C4B depending on 5 nucleotide differences in the C4d region that lead to differential biological functions (Yu et al 1986).
Altogether, over 41 C4 protein allotypes, determined by both common and specific nucleotide polymorphisms that lead to charge and serological differences in the plasma C4 proteins, have been identified (Mauff et al 1998). Physiologically, C4A has evolved to play crucial roles in the opsonization and immunoclearance processes, while C4B specializes in the efficient propagation of the classical and lectin pathways in a temporally and spatially specific manner (Carroll et al 1990, Dodds et al 1996). Both C4A and C4B exhibit unusually frequent variation in the gene number and a wide range of serum protein levels among different individuals and ethnic groups (100 -1000 µg/ml).

There may be one, two, three or four copies of functional C4 genes on each chromosome 6. It may code for a C4A or a C4B protein. Each C4 gene may be either 21 kb (long gene) or 14.6 kb (short gene) in size, due to the integration of the endogenous retrovirus HERV-K(C4) into intron 9 of the long gene (Dangel et al 1994). The C4 gene copy number per diploid genome, i.e. gene dosage, in the human population varies from 2 to 8. The variation of C4A and C4B gene dosage is always concurrent with the flanking genes RP, CYP21 and TNX in the MHC complement gene cluster. This RCCX (RP-C4-CYP21-TNX) modular variation is frequent in all populations (Blanchong et al 2000, Shen et al 1994, Yang et al 1999).

Correlation of the pathogenesis of SLE and C4 polymorphism was first noted in the Caucasian populations, as SLE patients carried a marked increase in the frequency of null alleles of C4, when compared with the normal population (Atkinson and Schneider 1999, Dawkins et al 1999, Fielder et al 1983). It was subsequently shown that the presence of C4A null alleles is a predisposing factor for SLE in different ethnic groups that include Caucasians, African Americans, Japanese, Korean and Chinese (Dunckley et
al 1987, Fielder et al 1983, Hawkins et al 1987, Hong et al 1994, Howard et al 1986, Olsen et al 1989, Steinsson et al 1998). The association of C4A deficiency with SLE is probably because C4A protein has a higher binding affinity to immune complexes, and to complement receptor 1 (CR1)(Birmingham et al 2001, Birmingham and Hebert 2001). Based on the observation that different polymorphic forms of complement C4 showed varying efficiencies of complement activation, it is suggested that susceptibility of C4 to infectious and autoimmune diseases would be related to the varying efficiencies of complement-mediated cell lysis and of immune aggregate dissolution by complement (Porter 1983). This in turn depends on the strength of interaction of different polymorphic forms of C4 with other proteins, such as C2, the C1 complex, C3, C5, the regulatory proteins complement receptor CR1, MBL complex, CRP and SAP, in the scheme of activation and inactivation of complement. An above average efficiency of complement activation might increase the risk of tissue damage during infections, while below average efficiency may slow the dissolution and removal of immune aggregates such as occurs in SLE.

However, previous investigations of C4 were largely based on phenotypic observations on the relative quantities of C4A and C4B proteins in the circulation, which do not accurately reflect their genetic compositions.

Complement C2 is about 36 kb away from C4 gene in the first locus in the MHC class III region, which is encoded by a 20 kb gene comprised of 18 exons. C2 is a single-chain glycoprotein and is the serine esterase component of the classical pathway C3 cleaving enzyme complex (Carroll et al 1984, Ishikawa et al 1990). A deficiency of C2 is a relatively frequent inherited defect of the complement system in individuals of western
European descent. In this population, approximately 1 person in 10,000 is homozygous C2 deficient (Glass et al 1976, Rynes et al 1982). C2 deficiency exhibits a strong linkage disequilibrium with certain HLA haplotypes. In humans, two variants of a genetic deficiency of C2 had been previously identified. Type I C2 deficiency is characterized by a 28 bp deletion across the 5’ splice junction of exon 6 that causes premature stop codon and no C2 protein translation. It is always associated with MHC haplotype A25 B18 C4A4 B2 DRw2, and occurs in over 90% of C2-deficient individuals (Johnson et al 1992). Type II C2 mutation occurs in 10% of C2-deficient individuals, where C2 secretion is impaired because of different missense mutations at highly conserved residues in each of the C2Q0 alleles. One is in exon 5 (nucleotide C566T; Ser189Phe) of the C2Q0 gene linked to the MHC haplotype A11 B35 C4AQ0 B1 DRw1. The other is in exon 11 (G1930A; Gly444Arg) of the C2Q0 gene linked to the MHC haplotype A2 B5 C4A3 B1 DRw4 (Wetsel et al 1996). About half of C2-deficient individuals have rheumatological disorders such as SLE, Henoch-Schonlein purpura, and polymyositis (Glass et al 1976). Similar to C4, C2 deficiency may lead to autoimmune diseases by insufficiency in complement activation and dissolution of immune complexes.

Located in HLA class II region, DR2 and DR3 had been consistently associated with SLE in many studied Caucasian populations with a two-fold relative risk conferred by each allele (Tsao 2002). Given the role of HLA class II molecules in the T-B cell interaction, it is not surprising that certain DR/DQ alleles are strongly associated with particular IgG autoantibody profiles. However, compared to Caucasians, the HLA DR association with SLE in non-Caucasians is less well-established (Tsao 2003).
TNFα, encoded by TNFA gene in MHC class III region, is a proinflammatory cytokine involved in the pathogenesis of many infectious and inflammatory diseases (Beutler and Cerami 1989). A polymorphism in TNFA promoter region –308 is associated with many autoimmune diseases including SLE. A significant increase was found in the frequency of TNFA –308A in lupus patients (Rood et al 2000). The paradox is that TNFA –308A is likely to increase the expression levels of TNFα, which was revealed to be protective against the pathogenesis of the lupus disease in animal experiments (Field 2001).

A comprehensive study of the susceptibility genes in MHC region will help understand the underlying mechanism of SLE disease development. However, definitive analyses of genetic factors in the MHC associated diseases have been complicated by the strong linkage disequilibrium of Class I, II and III alleles. For example, in one of the most common HLA haplotypes, HLA A1 Cw7 B8 DR3 (DRB1*0301 DQB*0201 DPB*0301), which is associated with many autoimmune diseases such as SLE and type I diabetes, the C4A gene is always deleted and the -308 position of the TNF-α promoter is always the A-nucleotide. As a result, it is difficult to distinguish whether it is a specific HLA class II allele, or the C4A gene deletion, or the -308A allele of TNF-α, or an uncharacterized mutation in one of the 128 expressed MHC genes, or a combination of all these factors that contribute to the disease pathogenesis.

To evaluate the effects of susceptibility genes in MHC region on SLE disease development, over 300 SLE patients, including Caucasians and African Americans, together with their unaffected family members were recruited. The variations of C4A and C4B gene dosages and RCCX length variants, C4A and C4B protein polymorphisms and
protein levels, C4 activation product C4a levels, and C4 mutation have been accurately determined using the well-established methods in our laboratory. In addition, the genetic variations in two other MHC class III genes including C2 and TNFA, were analyzed.
4.3 MATERIALS AND METHODS

4.3.1 Recruitment of human subjects and preparation of DNA and EDTA-plasma

The study populations included 229 Caucasian SLE patients (209 female and 20 male) and 56 African American SLE patients (53 female and 3 male). Available parents and unaffected siblings of all patients were also included to serve as internal controls. In most situations, at least two unaffected family members, either the two parents or one parent one or more siblings, were obtained. Patients and family members were recruited from the Nephrology Clinic and the Rheumatology Clinic of the Ohio State University Hospitals, the Arthritis Clinic and Nephrology Clinic of the Columbus Children’s Hospital and other Rheumatology Clinics in central Ohio. In addition, 150 female healthy Caucasians and 163 healthy African Americans who were Columbus Children’s hospital staff members or volunteers from local health fairs and churches were used as external controls. Peripheral blood samples were taken from Ohio SLE patients, unaffected family members and unrelated healthy controls upon informed consent, after the approval from the Columbus Children’s Hospital Institutional Review Board or the Ohio State University Institutional Review Board. Genomic DNAs were isolated from 3 ml of peripheral blood according to the standard protocol (Yu et al 2002). EDTA-plasma was separated from RBC and WBC by centrifugation of fresh peripheral blood.

4.3.2 Determination of the number and size of C4A and C4B genes and RCCX structures by TaqI and PshAI-PvuII RFLPs (Restriction Fragment Length Polymorphisms)

Genomic DNAs were digested by restriction enzymes, TaqI, PshAI and PvuII, and then analyzed by Southern blotting. In TaqI RFLP, three different probes, RP, CYP21A
and XA, were applied to the hybridization mixture simultaneously, and the band patterns of RP, C4, CYP21 and TNX were displayed clearly. Analyses of the presence and the relative band intensity in each DNA sample told us the number and size of C4 genes and RCCX structures on both chromosomes. The \textit{PshAI} RFLP allowed us to determine the relative dosages of C4A and C4B genes, as an additional \textit{PshAI} restriction site is present in C4A gene, but not in C4B gene. The addition of another enzyme, \textit{PvuII}, further reduced the restriction fragments in both C4A and C4B, allowing a clear assessment of the relative band intensities. The detailed technique approaches were discussed in our previous publications (Chung \textit{et al} 2002\textit{c}, Yu \textit{et al} 2002).

\textbf{4.3.3 Ascertainment of the RCCX haplotype and number and size of C4 genes on each chromosome 6 by \textit{PmeI} PFGE (Pulsed Field Gel Electrophoresis)}

To further confirm the presence of trimodular, bimodular and monomodular structures in the patients and related controls, high molecular weight genomic DNA embedded in low gelling temperature agarose were prepared from white cells of peripheral blood. The human genomic DNA, trapped in agarose block to avoid shearing and mechanical breakage, was digested with the rare cutter restriction enzyme \textit{PmeI}, subjected to PFGE and Southern blot analysis. The RCCX structures were represented as follows: Monomodular- 107 kb (S) or 113 kb (L); Bimodular- 139 kb (LS) or 146 kb (LL); Trimodular- 165 kb (LSS), 172 kb (LSL or LLS) or 179 kb (LLL). Results of the PFGE experiments allowed a clear depiction of the RCCX modular combinations. These two methods (RFLP-Southern blot and PFGE-Southern blot) helped us to elucidate the C4 genotype in each individual unambiguously.
4.3.4 Determination of C4A and C4B allotypes by C4 immunofixation

To correlate genomic C4 data with expressed C4 proteins, the allotypes of C4A and C4B were determined by high voltage agarose gel electrophoresis (HVAGE) of EDTA-plasma samples, followed by immunofixation using goat anti-human C4 antibody. The protein bands were observed by staining with Simple Blue dye (Invitrogen). Combining the C4 genotype with C4 allotypes from each subject allowed us to determine the C4 haplotype on each chromosome 6.

4.3.5 Determination of total C4 levels in plasma by radial immunodiffusion (RID)

RID was applied to determine total C4 levels in plasma using commercial kit (The Binding Site Ltd, Birmingham, UK). The concentrations of total C4 in each sample were calculated from a standard curve based on known concentrations of C4 protein.

4.3.6 Determination of C4A and C4B concentrations

To determine the levels of C4A and C4B proteins in plasma, C4A and C4B allotypes from each individual were scanned from the allotyping gels using an Epson Expression 1600 Scanner, and quantitated by densitometry using ImageQuant Version 5.0 software. The C4A and C4B plasma levels were then deduced according to total C4 protein levels.

4.3.7 Detection of plasma C4 split product C4a levels by ELISA

ELISA was applied to detect the plasma levels of complement C4 activation product C4a using a commercial kit (PharMingen, San Diego, CA). The titer of plasma C4a in each sample was calculated based on the standard curve obtained from standards with known concentrations.
4.3.8 Detection of known C4 and C2 mutations in a single multiplex PCR reaction

To rapidly screen genomic DNA samples for known C4 and C2 mutations, a multiplex PCR strategy was designed. The sequence specific primers covering the C4 mutation (2 bp insertion in exon 29 of C4 gene) and type I C2 mutation (28 bp deletion in exon 6 of C2 gene) were used to amplify regions in C4 and C2 genes. The primer sequences were as follows: C4E26.5 (5’GCT CAC AGC CTT TGT GTT GAA3’), C4E29insR (5’GAG AAC CAG TGA CTG AGA GC3’), 5C2 (5’GCC TGG GCC GTA AAA TCC AAA TCC A3’), 3C2 (5’GCA CAG GAA GGC CTC TGC TGC AGG C3’).

The PCR condition was 94°C 3 minutes; 94°C 30 seconds, 60°C 45 seconds, 72°C 1 minute for 30 cycles; 72°C 10 minutes, and finally 4°C for storage. An 174-bp band present in all samples indicated the presence of a normal C2 gene. If another 146-bp band was detected, the type I C2 mutation was present in that individual. The presence of an additional 868-bp band indicated a C4 mutation due to a 2 bp insertion exon 29. The products were separated by 1.5-2% agarose gel, and then stained by ethidium bromide.

4.3.9 Determination of TNFA -308 polymorphism by SSP-PCR (Sequence specific-primer-polymerase chain reaction)

The TNFA -308 polymorphisms were determined by SSP-PCR using primers TNF308F (5’GAG GCA ATA GGT TTT GAG GGC CAT3’) and TNF308R (5’CCC TGT GTG TTC GTA GTT C3’). The PCR conditions were 94°C 3 minutes; 94°C 30 seconds, 59°C 60 seconds, 72°C 2 minutes for 35 cycles; 72°C 5 minutes, and then 4°C for storage. The PCR products were digested by NcoI at 37°C, and then separated by 1.5% agarose gel. The G and A alleles were 20-bp apart after ethidium bromide staining.
4.3.10 Statistical analyses

*Chi-squared* analysis was performed to determine the differences in C4, C4A and C4B gene dosages, RCCX structures, C4 and C2 mutations and TNFA –308 polymorphisms among all groups. Univariate analysis of variance test (two-way ANOVA) was used to evaluate possible factors contributing to the normally distributed C4, C4A and C4B plasma protein levels. Tests were two-tailed. Statistical analysis was performed by SPSS 11.5 (SPSS, Chicago, IL) software.

4.3.11 Additional declarations

This is a major study going on in our laboratory. All the lupus patients and family members were recruited by Mrs. Maddie Hebert. Many students and staff contributed to produce the data presented in this chapter. Briefly, *TaqI* and *PshAI-PvuII* Southern blot analyses were performed by our research associate Bi Zhou. *PmeI* PFGEs were performed by Dr. Erwin Chung and our laboratory staff, Yaoling Shu.

The C4 protein levels were determined by the author using radial immunodiffusion with the help of Yaoling Shu. The C4 allotypes were determined by Bi Zhou using C4 immunofixation. The C4 allotyping gels were scanned and quantitated by the author, and then used to calculate the C4A and C4B protein levels. The C4a was determined by the author by ELISA. The SSP-PCRs for C4 and C2 mutations and TNFA -308 polymorphisms were designed by the author, finished by the author with the help of an honored undergraduate student Nora Alghothani, and two graduate students Jenny Kohout and Yee ling Wu. The data were then interpreted and put into our database by the author. Finally, statistical analyses were performed by the author with the guidance of our statistician, Dr. John Hayes.
4.4 RESULTS

4.4.1 The differences in C4 polygenic and RCCX modular variations between female healthy Caucasians and Caucasian SLE families

The study population included 209 female Caucasian SLE patients and at least two unaffected family members from every patient family. The C4 genetics in 150 Ohio female normal Caucasians, previously determined by Dr. Carol Blanchong in our laboratory (Blanchong et al 2000), was used as a reference for comparison. The genotypic diversities of the constituents of the RCCX modules, i.e., RP1 and RP2, long and short genes for C4A and C4B, CYP21A and CYP21B, and TNXA and TNXB in all Caucasians were determined by three RFLPs coupled with different restriction enzyme digestion. TaqI RFLP yield information about the total number of C4 long and short genes by the presence and relative dosages of RP1-C4 long (L), RP1-C4 short (S), RP2-C4 long (L) and RP2-C4 short (S) bands using a probe hybridized to RP gene. The presence and relative dosage of CYP21A and CYP21B, and TNXA and TNXB were detected in TaqI RFLP using two probes hybridized to CYP21 and TNX genes. The number of C4A and C4B genes in each individual was determined by PshAI-PvuII RFLP using a C4d probe. Furthermore, Pmel PFGE was used to confirm the number and size of C4 genes and RCCX module on each chromosome 6.

When we compared SLE patients with unrelated normal controls, SLE patients had significantly increased frequencies of low C4 and C4A gene dosages than controls (Figure 4.1 and Table 4.1). The frequency of subjects with two or three C4 genes was 41.1% in SLE patients, but only 27.3% in unrelated controls. On the contrary, individuals with five or six C4 genes were only 4.8% in SLE, compared with 20.6% in controls.
There was an almost equal frequency of individuals with four C4 genes between the patient and the control populations. This difference in C4 gene dosage between SLE and normals was highly significant, with a $p$ value of 0.000018 using chi-squared analysis. The same phenomenon was observed in C4A gene dosages. The frequency of SLE patients with zero or one C4A gene was 32.2%, which was significantly higher than 20.7% in controls ($p=0.00041$). Unlike total C4 and C4A, the majority of SLE patients had two C4B genes. There was also a significant difference between SLE and controls for C4B genes ($p=0.0006$). However, this was clearly due to the different distribution in the frequencies. It was noticed that the average C4B gene dosage per individual in SLE patients (1.740) was slightly lower than those in controls (1.773). Consistent with the increased frequencies of low C4 and C4A genes, the average C4 and C4A gene dosages in SLE were 3.569 and 1.832, respectively, which were lower than 3.947 and 2.167, respectively, in controls. RCCX modular patterns between SLE and controls were highly different ($p=0.0000066$). SLE patients had higher frequencies of mono-S and LS, but lower frequencies of LL, LSS and LLL than normal controls. The mono-S (i.e. S) RCCX haplotype increased from 11% in controls to 19.1% in SLE patients (haplotype frequency). The ratio of long and short C4 genes was only 2.545 in SLE, but 3.191 in controls indicating the overall frequency of short C4 genes was increased in SLE.

Intriguingly, when we examined the SLE patients’ mothers and female siblings, similar trends were observed. SLE mothers and siblings had also reduced dosages for total C4 and C4A genes, although the deficiency was more severe in the mothers than in the siblings. This is understandable that considering SLE is a complex disease, and many genes are involved in the disease manifestation.
4.4.2 The variation in total C4, C4A and C4B protein levels in female Caucasian SLE patients and unaffected family members

The total C4 plasma protein levels in SLE patients and unaffected family members were determined by radial immunodiffusion. The ratios of C4A and C4B allotypes were calculated by quantitating the C4A and C4B protein bands from allotyping gel. The levels of C4A and C4B proteins were then deduced from total C4 protein levels based on the ratio of band intensities. It was observed that the protein levels of total C4, C4A and C4B in SLE patients were consistently lower than those in the mothers and female siblings (Figure 4.2). This phenomenon was more apparent for total C4 (p<0.001) and C4A (p<0.001) than for C4B (p=0.038). The higher frequencies of low total C4 and C4A gene dosages in the SLE patients may explain these results. The female siblings had almost the same levels of total C4, C4A and C4B proteins as the mothers.

In our previous study with a group of Hungarian Caucasians, it was noticed that total C4, C4A and C4B protein levels were all positively correlated with the corresponding gene dosages (Yang et al. 2003). In this study, we found that there were also positive correlations between gene dosages of C4A and C4B and the related protein levels for all three groups, including SLE patients (p<0.001, \(R^2=0.395\) for C4A; \(p<0.001, R^2=0.242\) for C4B), mothers (p<0.001, \(R^2=0.297\) for C4A; \(p<0.001, R^2=0.464\) for C4B) and siblings (p<0.001, \(R^2=0.325\) for C4A; \(p<0.001, R^2=0.421\) for C4B) (Figure 4.3). The correlations were highly significant with \(R^2\) varying from 0.242 to 0.464. Total C4 protein levels in SLE patients (p=0.032, \(R^2=0.042\)) and mothers (p<0.001, \(R^2=0.229\)) were significantly correlated with C4 gene dosages, but to a less extent. However, there
was no clear correlation between C4 protein levels and C4 gene dosages in the female siblings.

A relevant phenomenon observed in this study was that SLE patients had relatively lower total C4, C4A and C4B protein levels than their mothers and female siblings, even within the same gene dosage group (Figure 4.3). The lower protein levels in SLE patients were more obvious in total C4 and C4A than in C4B. This was consistent with the lower total C4 and C4A gene dosages in SLE patients. Possible explanations of the lower C4 protein levels included C4 consumption induced by complement activation or inflammatory response, low or no C4 expression caused by C4 genetic variations, or both.

In conclusion, the protein levels of total C4, C4A and C4B were reduced in the SLE patients by lower gene dosages and probably, some other genetic and inflammatory factors.

4.4.3 C4 mutations in Caucasian SLE

A common mutation in exon 29 of the C4 gene that could lead to premature stop codon and no C4 expression was screened in the Caucasian SLE patients and unrelated controls using PCR strategy. In total, 190 female Caucasian SLE patients and 150 unrelated controls were included. This mutation was found in eleven SLE patients. PshAI Southern blot analysis of PCR products indicated all mutated C4 genes were C4A genes. Further analysis suggested all C4A mutations were from a bimodular LS RCCX haplotype associated with C4B2. In controls, the same mutation was only found in three individuals. Although it was not significant using chi-squared analysis (p=0.081), the mutation rate increased from 2% in controls to 5.79% in SLE patients.
4.4.4 The effect of sex on C4 genetic variation and C4 protein expression in SLE patients and unaffected family members

Since SLE is an autoimmune disease that mainly affects females (female: male=9:1), it is of interest to investigate the role of sex in the disease manifestation. In this study, we compared the C4 genetic variations and protein levels between female and male SLE patients. It was noticed that there were no significant differences in the variations of total C4, C4A and C4B gene dosages and RCCX modular structures between female and male SLE patients ($p>0.50$) (Table 4.2). However, some minor differences were still observed. The frequency of C4A deficiency in male SLE was not as high as that in female SLE. Although male SLE patients had slightly higher frequency of three C4 genes, the majority of them had two C4A and two C4B genes. Same as female SLE, monomodular Short (S) was increased in male SLE.

It was unexpected that the protein levels of total C4 ($p=0.042$), C4A ($p=0.041$) and C4B ($p=0.107$) in female SLE were lower than those in male SLE (Figure 4.4). This could not be explained by the differences in the gene dosages of total C4, C4A and C4B. Unlike SLE patients, the protein levels of total C4, C4A and C4B in unaffected female family members (including mothers and female siblings) were a little higher than male family members (including father and male siblings) ($p>0.05$) (Figure 4.5). A possible reason would be the difference in the effect of female sex hormone on C4 protein expression during disease and normal conditions.

4.4.5 The effect of other susceptibility genes in MHC region on Caucasian SLE

In addition to complement C4 gene, two other genes in the MHC class III region had been reported to be associated with SLE development. One is complement C2,
complement gene immediately downstream of C4 in the complement activation pathways. The other is TNFA gene that expresses a proinflammatory cytokine involved in several autoimmune diseases such as rheumatoid arthritis (Krause et al 2003, Sfikakis and Kollias 2003). In order to evaluate the role of these two factors in the SLE development, type I complement C2 mutation and TNFA -308 polymorphisms that could regulate TNFA expression had been screened in this Caucasian SLE population. C2 mutation was detected in a PCR reaction together with known C4 mutation. The TNFA –308 polymorphism was determined by an SSP-PCR strategy coupled with a restriction enzyme NcoI digestion.

Although C2 seems to be associated with SLE in some studies, in our Caucasian SLE population (n=190), only three patients had this type I C2 mutation. This suggested C2 could be a secondary factor for SLE development in our Caucasian SLE population.

The TNFA -308A allele in Caucasians was strongly associated with complement component C4A deficiency, particularly in individuals with the HLA haplotype A1, B8, DRB1*0301, DQA1*0501, and DPB1*0201. This haplotype was also known as the ancestral haplotype (AH) 8.1 (Candore et al 2002, Degli-Esposti et al 1992, Price et al 1999, Wilson et al 1993). It was reported that the AH-8.1 had a short C4B1 gene from a monomodular RCCX structure (Carroll et al 1985, Chung et al 2002b, Dangel et al 1994, Ulgiati et al 1996). In 116 Caucasian SLE patients, 45.7% had this A allele at –308, while subjects with the same allele in controls were only 32.5% (Table 4.3). This difference was modestly significant, with a p value of 0.036 using chi-squared analysis. To further analyze the data, the association of the TNFA –308 A allele with monomodular short (i.e. mono-S) RCCX haplotype was studied. Five out of six SLE patients with homozygous A
alleles were associated with mono-S on both chromosomes. 57.7% to 58.5% TNFA –308 A allele carriers were associated with mono-S in both normal controls and SLE patients, but 30.2% to 34.6% carriers were clearly not associated with mono-S. Notably, the –308A carriers with mono-S increased from 18.8% in controls to 26.7% in SLE patients, while –308A carriers without mono-S increased very little. This suggested that the higher frequency of TNFA –308 A allele in SLE was mainly from the association with mono-S. In other words, this A allele could be an indicator, but not the primary risk factor in SLE disease.

4.4.6 The comparison of C4 polygenic and RCCX modular variations and C4 protein expression between normal African Americans and African American SLE patients

Besides Caucasian SLE, a small African American SLE population (N=56) was studied. The variations in the C4, C4A and C4B gene dosages and RCCX structures in African American SLE were determined using the same techniques and then compared with normal African Americans (N=163) described in chapters 3 and 4. It was observed that the low dosages of total C4 and C4A were increased in African American SLE. The frequencies of individuals with two or three C4 genes increased from 18.4% in normal African Americans to 33.9% in African American SLE, although it was not significant ($p=0.136$) (Table 4.4). Homozygous or heterozygous deficiencies of C4A (that is, C4A gene dosage equals to 0 or 1) increased from 8% in normal controls to 16% in SLE patients. Mono-S increased from 4.6% in controls to 8.2% in SLE. The deficiencies in total C4 and C4A genes were further reflected by the mean C4 and C4A gene dosages (i.e. gene index). C4 gene index decreased from 3.957 in controls to 3.750 in SLE, while C4A gene index decreased from 2.129 to 2.000. No difference was observed for the ratio of
long and short C4 gene between patient and control groups. Although the sample size limits the significance of this study, the same trend in African American and Caucasian SLE seemed to suggest the role of C4, especially C4A deficiency, in SLE development.

In Figure 4.6, we compared the average protein levels of total C4, C4A and C4B between normal African Americans and African American SLE patients. Consistent with the gene dosage variations, the protein levels of total C4 \( p=0.004 \) and C4A \( p=0.036 \) in SLE were lower than those in normal controls. C4B protein levels in SLE were slightly lower than that in controls, but it was not statistically significant \( p=0.068 \).

4.4.7 The complement C4 activation in SLE patients

In the circulation, complement C4 can be activated through C1s and MASP2 in complement activation pathways. Complement pathways can be initiated during infections and inflammations. In autoimmune diseases such as SLE, the immune complex, comprised of autoantibodies and autoantigens, could trigger classical complement activation pathway and C4 activation. The C4a, the split product after C4 activation, had been used as an indicator for SLE disease activity in many studies. In this study, plasma C4a levels had been determined in 74 Caucasian SLE and 27 African American SLE using C4a ELISA. As expected, both African American and Caucasian SLE patients had increased plasma levels of C4a. No significant difference in C4a plasma levels was observed between Caucasian and African American SLE patients \( p=0.982 \) (Figure 4.8).

The relationship of total C4, C4A and C4B protein levels with plasma C4a levels were studied and shown in Figure 4.9. There were no clear correlations observed between protein levels of total C4, C4A and C4B and plasma C4a levels in SLE. Most SLE patients had low levels of complement C4 activation, regardless of the protein levels of
C4, C4A and C4B. A small portion of patients had high plasma C4a levels and relatively low protein levels of total C4, C4A and C4B, indicating complement activation or disease relapse. The consistently higher C4a levels in SLE patients, even during remission, probably indicated the undergoing pathological changes and disease progression in SLE, suggesting the essentiality of treatment during remissions.

4.4.8 A comparison of C4 polygenic and RCCX modular variations and C4 protein expression between Caucasian and African American SLE patients

It was observed for a long time that African American and Caucasian SLE had their own characteristics. African Americans tended to have more severe SLE disease and more clinical complications such as lupus nephritis than Caucasians (Wallace and Hahn 1993). The underlying mechanism was not clear. In our study, we compared the C4 genetic variations and protein polymorphisms between Caucasian and African American SLE. The results were summarized in Table 4.5 and Figures 4.7-4.8.

No significant results were obtained for the variations in total C4 \( (p=0.2995) \), C4A \( (p=0.0846) \) and C4B \( (p=0.5651) \) gene dosages between two SLE groups. For RCCX modular variations, the African American SLE patients had significantly increased frequency of LS RCCX haplotype and decreased LL haplotype than Caucasian SLE. The ratio of long and short C4 gene ranged from 1.315 in African American SLE to 2.540 in Caucasian SLE. Although both African American and Caucasian SLE were deficient in total C4 and C4A genes, it was noticed that the deficiencies were more obvious in Caucasian SLE. The C4 gene index was 3.750 in African American SLE, but only 3.569 in Caucasian SLE. The C4A gene index decreased from 2.000 in African American SLE to 1.809 in Caucasian SLE.
The average protein levels of total C4 and C4A somehow reflected the differences in total C4 and C4A gene index between these two SLE groups. African American SLE tended to have relatively higher total C4 and C4A protein levels than Caucasian SLE. The protein levels of C4B between two groups were almost the same. It was also noticed, however, that the differences in protein levels were not statistically significant using two-way ANOVA.
4.5 DISCUSSION

In this study, the variations in total C4, C4A and C4B gene dosages and RCCX structures; C4 polymorphisms; C4, C4A and C4B protein levels and C4 mutations were determined in 229 Caucasian SLE patients, 56 African American SLE patients and their unaffected family members. C4 activation product C4a levels were determined in some Caucasian and African American SLE patients. In addition, two other susceptibility genes including TNFA and C2 were screened in the majority of Caucasian SLE patients for TNFA –308 polymorphism and type I C2 mutation.

The role of complement C4 in SLE

It was observed about 50 years ago that low serum complement activity or low protein concentrations of complement C4 concurred with disease activities of SLE. However, most past epidemiological studies of C4 in human SLE did not consider the polygenic and gene size variations of C4A and C4B. In addition, many studies were overly dependent on phenotypic observations or methods that did not distinguish differential C4A and C4B protein expression caused by unequal gene number or different gene size from the absence of a functional C4A and C4B gene. An accurate account of polygenic variation of C4A and C4B has never been performed. Determination of C4A and C4B gene dosages, C4 protein polymorphisms and protein levels in SLE patients allowed us to examine the effect of C4 genetic deficiencies, over-dosages and/or low and over expression of C4 in SLE more precisely. An anticipated difficulty in the study of SLE and complement C4 is the linkage disequilibrium of specific RCCX haplotypes with specific alleles located in the MHC region. A typical example is AH8.1, which includes
mono-S RCCX haplotype with C4A deficiency that is always linked to TNF-308 A allele and HLA DR3.

In our lupus study with Caucasian SLE from central Ohio, it was noticed that there was a highly significant difference in C4 \((p=0.000018)\) and C4A \((p=0.00041)\) gene dosages between SLE and unrelated controls from the same region. The difference was caused by the shift of the distribution to the left in the SLE patient group, indicating SLE patients had higher frequencies of low C4 and C4A gene dosages than controls. The lower C4 \((3.569 \text{ vs } 3.947)\) and C4A \((1.832 \text{ vs } 2.167)\) gene indeces in SLE than in controls further substantiated this hypothesis. Different from total C4 and C4A genes, the mean C4B gene dosages between patients and controls \((1.740 \text{ vs } 1.773)\) were very close, although the distribution of C4B gene dosage variation was different \((p=0.0006)\). Therefore, deficiency of C4 and C4A genes but not C4B is associated with SLE in Ohio Caucasians. Another dimension of C4 genetic difference between patients and controls was caused by the RCCX modular variation \((p=0.0000066)\). SLE patients had higher frequency of LS and mono-S, but lower frequency of LL, LSS and LLL than controls, which led to lower ratios of long and short C4 genes from 3.191 in controls to 2.545 in SLE. It was reported in our previous study that short C4 gene made more C4 protein than long C4 gene (Yang et al 2003). This result may suggest some SLE patients having normal or high C4 gene dosages with short C4 genes may express relatively high C4 protein levels and help aggravate the disease during flares. Detailed medical information assisting in segregating patients based on clinical symptoms, age of disease onset, disease severity and response to therapies may help understanding the role of C4 in the disease pathophysiology. Intriguingly, SLE patients’ mother and siblings had the similar molecular defect in total C4 and C4A.
Consistent with the defect in C4 and C4A genes, the average protein levels of total C4 and C4A in SLE patients were lower than those in mothers and siblings. There was a positive correlation between total C4, C4A and C4B gene dosages and the corresponding protein levels in patients, mother and female sibling. Strikingly, the protein levels of C4, C4A and C4B in SLE patients were consistently lower than those in mothers and siblings, even within the same gene dosage group. Possible explanations include C4 mutation that leads to no C4 expression; low C4 expression that is caused by low C4 promoter activity or suppression by some factors fluctuated during the disease; and C4 consumption during secondary infection and inflammation.

In order to confirm our hypothesis for the association of C4 and SLE, a African American SLE population (N=56) was recruited and studied with 163 unrelated normal controls. Similar to Caucasians, African American SLE had higher frequencies of low C4 and C4A gene dosages than controls, although it had not reached statistical significance. Consistent with this result, African American SLE also had relatively low mean C4 ($p=0.004$) and C4A ($p=0.036$) protein levels than controls. The same phenomenon, reduce dosages in total C4 and C4A genes and accompanying low C4 and C4A plasma protein levels in both Caucasian and African American SLE patients suggest the role of C4, especially C4A deficiency, in SLE pathogenesis.

Conclusions

SLE is a multifactorial disease that both genetic and environmental factors are involved in disease development. Our study in the polygenic and gene size variations of C4A and C4B, the C4A and C4B protein expression and protein polymorphisms and split product C4a levels from Caucasian and African American SLE suggested C4 is an
important factor in the SLE development. Deficiencies of total C4 and C4A in both Caucasian and African American SLE cannot be explained in full by the linkage disequilibrium of C4A deficiency with mono-S (18%). The increased frequency of C4A mutation in SLE (5.79%) helped further strengthened our hypothesis. Strikingly, the levels of total C4, C4A and C4B proteins in the circulation of SLE patients were lower than those in the unaffected family members, even within the same gene dosage group. Overall, the C4 and C4A genetic deficiencies, C4A mutations, and lower C4, C4A and C4B expression may all decrease the plasma C4, especially C4A protein levels in the patients, which probably predispose subjects to disease development. Other factors in MHC region, including C2, TNFA and DR, seems to play a minor role in SLE development based on our studies.
Figure 4.1. A comparison of the variations in gene dosages of total C4 (panel A), C4A (panel B), C4B (panel C) and RCCX structures (panel D) between normal and SLE Caucasians. Dark blue bars represented normal female Caucasians (N=150), while red bars represented female Caucasian SLE (N=209). P values were indicated based on chi-squared analysis.
Figure 4.2 A comparison of total C4, C4A and C4B protein levels among female Caucasian SLE patients (N=209), mothers (N=164) and siblings (N=110). The p value and R-squared were indicated based on two-way ANOVA.
Figure 4.3 Correlations of total C4, C4A and C4B protein levels with the corresponding gene dosages in female SLE patients (red bars, N=209), mothers (yellow bars, N=164) and female siblings (green bars, N=110). It was noticed that total C4, C4A and C4B protein levels were all positively correlated with C4, C4A and C4B gene dosages within each group. Intriguingly, the protein levels of C4, C4A and C4B in SLE patients were consistently lower than those in the mothers and siblings even within the same gene dosage group.
Figure 4.4 A comparison of the mean total C4, C4A and C4B protein levels between female and male Caucasian SLE patients. The $p$ values and R-squared were calculated based on univariate analysis of variance.
Figure 4.5 A comparison of the mean total C4, C4A and C4B protein levels among female and male Caucasian SLE family members. Panels A, C and E showed the difference in protein levels between SLE mother (N=160) and SLE father (N=101). Panels B, D and F showed the difference in protein levels between female (N=108) and male siblings (N=61). No significant differences were observed based on analysis of variance.
Figure 4.6 A comparison of the total C4, C4A and C4B protein levels among normal African Americans (N=162) and African American SLE patients (N=54). The $p$ values were indicated based on two-way ANOVA.
Figure 4.7 A comparison of the total C4, C4A and C4B protein levels among African American (N=54) and Caucasian SLE patients (N=236). No significant results were observed based on two-way ANOVA.
Figure 4.8 A comparison of complement C4 activation product C4a in the circulation between African American (N=27) and Caucasian (N=74) SLE patients. The p value was indicated based on analysis of variance.
Figure 4.9 The effect of total C4, C4A and C4B protein levels on the plasma levels of complement C4 activation product C4a. No clear correlations were observed.
<table>
<thead>
<tr>
<th>Number of C4 genes</th>
<th>Female Caucasians</th>
<th>Caucasian Female SLE</th>
<th>SLE Mother</th>
<th>Female Sibling</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3</td>
<td>0.020</td>
<td>14</td>
<td>0.067</td>
</tr>
<tr>
<td>3</td>
<td>38</td>
<td>0.253</td>
<td>72</td>
<td>0.344</td>
</tr>
<tr>
<td>4</td>
<td>78</td>
<td>0.520</td>
<td>113</td>
<td>0.541</td>
</tr>
<tr>
<td>5</td>
<td>26</td>
<td>0.173</td>
<td>10</td>
<td>0.048</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>0.033</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>Total</td>
<td>150</td>
<td>209</td>
<td>164</td>
<td>110</td>
</tr>
</tbody>
</table>

Chi-square test (vs Normal) $p = 0.000018$ $p = 0.0078$ $p = 0.1375$

<table>
<thead>
<tr>
<th>Number of C4A genes</th>
<th>Female Caucasians</th>
<th>Caucasian Female SLE</th>
<th>SLE Mother</th>
<th>Female Sibling</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0.007</td>
<td>11</td>
<td>0.053</td>
</tr>
<tr>
<td>1</td>
<td>30</td>
<td>0.200</td>
<td>56</td>
<td>0.269</td>
</tr>
<tr>
<td>2</td>
<td>69</td>
<td>0.460</td>
<td>108</td>
<td>0.519</td>
</tr>
<tr>
<td>3</td>
<td>44</td>
<td>0.293</td>
<td>24</td>
<td>0.115</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>0.033</td>
<td>8</td>
<td>0.038</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.007</td>
<td>1</td>
<td>0.005</td>
</tr>
<tr>
<td>Total</td>
<td>150</td>
<td>208</td>
<td>164</td>
<td>108</td>
</tr>
</tbody>
</table>

Chi-square test (vs Normal) $p = 0.00041$ $p = 0.0385$ $p = 0.0689$

<table>
<thead>
<tr>
<th>Number of C4B genes</th>
<th>Female Caucasians</th>
<th>Caucasian Female SLE</th>
<th>SLE Mother</th>
<th>Female Sibling</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3</td>
<td>0.020</td>
<td>8</td>
<td>0.038</td>
</tr>
<tr>
<td>1</td>
<td>48</td>
<td>0.320</td>
<td>45</td>
<td>0.216</td>
</tr>
<tr>
<td>2</td>
<td>80</td>
<td>0.533</td>
<td>148</td>
<td>0.712</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>0.120</td>
<td>7</td>
<td>0.034</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0.007</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>Total</td>
<td>150</td>
<td>208</td>
<td>164</td>
<td>108</td>
</tr>
</tbody>
</table>

Chi-square test (vs Normal) $p = 0.0006$ $p = 0.0024$ $p = 0.0061$

<table>
<thead>
<tr>
<th>RCCX haplotype</th>
<th>Female Caucasians</th>
<th>Caucasian Female SLE</th>
<th>SLE Mother</th>
<th>Female Sibling</th>
</tr>
</thead>
<tbody>
<tr>
<td>LL</td>
<td>138</td>
<td>0.460</td>
<td>183</td>
<td>0.438</td>
</tr>
<tr>
<td>LS</td>
<td>69</td>
<td>0.230</td>
<td>111</td>
<td>0.266</td>
</tr>
<tr>
<td>S</td>
<td>33</td>
<td>0.110</td>
<td>80</td>
<td>0.191</td>
</tr>
<tr>
<td>L</td>
<td>18</td>
<td>0.060</td>
<td>26</td>
<td>0.062</td>
</tr>
<tr>
<td>LSL</td>
<td>1</td>
<td>0.003</td>
<td>3</td>
<td>0.007</td>
</tr>
<tr>
<td>LSS</td>
<td>19</td>
<td>0.063</td>
<td>6</td>
<td>0.014</td>
</tr>
<tr>
<td>LLL</td>
<td>22</td>
<td>0.073</td>
<td>7</td>
<td>0.017</td>
</tr>
<tr>
<td>Others(LSSS &amp; SS)</td>
<td>0</td>
<td>0.000</td>
<td>2</td>
<td>0.005</td>
</tr>
<tr>
<td>Total</td>
<td>300</td>
<td>418</td>
<td>328</td>
<td>220</td>
</tr>
</tbody>
</table>

Chi-square test (vs Normal) $p = 0.000006$ $p = 0.0021$ $p = 0.0047$

<table>
<thead>
<tr>
<th>Number of L genes</th>
<th>Female Caucasians</th>
<th>Caucasian Female SLE</th>
<th>SLE Mother</th>
<th>Female Sibling</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>450</td>
<td>0.761</td>
<td>537</td>
<td>0.718</td>
</tr>
<tr>
<td>Number of S genes</td>
<td>141</td>
<td>0.239</td>
<td>211</td>
<td>0.282</td>
</tr>
<tr>
<td>C4 gene index</td>
<td>3.947</td>
<td>3.569</td>
<td>3.665</td>
<td>3.745</td>
</tr>
<tr>
<td>C4B gene index</td>
<td>1.773</td>
<td>1.740</td>
<td>1.701</td>
<td>1.694</td>
</tr>
<tr>
<td>C4A gene: C4B gene</td>
<td>1.222</td>
<td>1.053</td>
<td>1.154</td>
<td>1.208</td>
</tr>
</tbody>
</table>

Table 4.1 The variations in C4, C4A and C4B gene dosages and RCCX modular structures among female normal Caucasians, female Caucasian SLE patients, mothers and siblings.
<table>
<thead>
<tr>
<th></th>
<th>Caucasian Female SLE</th>
<th>Caucasian Male SLE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Freq.</td>
</tr>
<tr>
<td><strong>Number of C4 genes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>0.067</td>
</tr>
<tr>
<td>3</td>
<td>72</td>
<td>0.344</td>
</tr>
<tr>
<td>4</td>
<td>113</td>
<td>0.541</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>0.048</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>209</td>
<td></td>
</tr>
<tr>
<td><em>Chi-square test</em></td>
<td>p=0.960</td>
<td></td>
</tr>
<tr>
<td><strong>Number of C4A genes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>11</td>
<td>0.053</td>
</tr>
<tr>
<td>1</td>
<td>56</td>
<td>0.269</td>
</tr>
<tr>
<td>2</td>
<td>108</td>
<td>0.519</td>
</tr>
<tr>
<td>3</td>
<td>24</td>
<td>0.115</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>0.038</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.005</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>208</td>
<td></td>
</tr>
<tr>
<td><em>Chi-square test</em></td>
<td>p=0.696</td>
<td></td>
</tr>
<tr>
<td><strong>Number of C4B genes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>8</td>
<td>0.038</td>
</tr>
<tr>
<td>1</td>
<td>45</td>
<td>0.216</td>
</tr>
<tr>
<td>2</td>
<td>148</td>
<td>0.712</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>0.034</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>208</td>
<td></td>
</tr>
<tr>
<td><em>Chi-square test</em></td>
<td>p=0.7998</td>
<td></td>
</tr>
<tr>
<td><strong>RCCX haplotype</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LL</td>
<td>183</td>
<td>0.438</td>
</tr>
<tr>
<td>LS</td>
<td>111</td>
<td>0.266</td>
</tr>
<tr>
<td>S</td>
<td>80</td>
<td>0.191</td>
</tr>
<tr>
<td>L</td>
<td>26</td>
<td>0.062</td>
</tr>
<tr>
<td>LSL</td>
<td>3</td>
<td>0.007</td>
</tr>
<tr>
<td>LSS</td>
<td>6</td>
<td>0.014</td>
</tr>
<tr>
<td>LLL</td>
<td>7</td>
<td>0.017</td>
</tr>
<tr>
<td>Others(LLSS &amp; SS)</td>
<td>2</td>
<td>0.005</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>418</td>
<td></td>
</tr>
<tr>
<td><em>Chi-square test</em></td>
<td>p=0.6478</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.2 The differences in the variations of C4, C4A and C4B gene dosages and RCCX modular structures between female and male Caucasian SLE patients.
Table 4.3 The polymorphism of TNFA at -308 in Caucasian SLE and unrelated normal controls.

<table>
<thead>
<tr>
<th></th>
<th>SLE (n =116)</th>
<th>Controls (n = 80)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>homozygous carriers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mono-S/mono-S</td>
<td>5 (4.3%)</td>
<td>0 (0.00%)</td>
</tr>
<tr>
<td>mono-S/LL</td>
<td>1 (0.9%)</td>
<td>1 (1.25%)</td>
</tr>
<tr>
<td>mono-S/LSS</td>
<td>0 (0.0%)</td>
<td>1 (1.25%)</td>
</tr>
<tr>
<td><strong>heterozygous carriers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>with mono-S</td>
<td>31 (26.7%)</td>
<td>15 (18.8%)</td>
</tr>
<tr>
<td>without mono-S</td>
<td>16 (13.8%)</td>
<td>9 (11.3%)</td>
</tr>
<tr>
<td><strong>total carriers</strong></td>
<td>53 (45.7%)</td>
<td>26 (32.5%)</td>
</tr>
</tbody>
</table>

* $\chi^2 = 4.38; P = 0.036$

Table 4.3 The polymorphism of TNFA at -308 in Caucasian SLE and unrelated normal controls.
<table>
<thead>
<tr>
<th>Number of C4 genes</th>
<th>Normal Blacks</th>
<th>Black SLE</th>
<th>Chi-square test</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Freq.</td>
<td>N</td>
<td>Freq.</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>0.018</td>
<td>1</td>
<td>0.018</td>
</tr>
<tr>
<td>3</td>
<td>27</td>
<td>0.166</td>
<td>18</td>
<td>0.321</td>
</tr>
<tr>
<td>4</td>
<td>110</td>
<td>0.675</td>
<td>31</td>
<td>0.554</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>0.123</td>
<td>6</td>
<td>0.107</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>0.018</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>Total</td>
<td>163</td>
<td></td>
<td>56</td>
<td></td>
</tr>
</tbody>
</table>

Chi-square test $p = 0.136$

<table>
<thead>
<tr>
<th>Number of C4A genes</th>
<th>Normal Blacks</th>
<th>Black SLE</th>
<th>Chi-square test</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Freq.</td>
<td>N</td>
<td>Freq.</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0.000</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>1</td>
<td>13</td>
<td>0.080</td>
<td>9</td>
<td>0.161</td>
</tr>
<tr>
<td>2</td>
<td>121</td>
<td>0.742</td>
<td>40</td>
<td>0.714</td>
</tr>
<tr>
<td>3</td>
<td>24</td>
<td>0.147</td>
<td>5</td>
<td>0.089</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>0.031</td>
<td>2</td>
<td>0.036</td>
</tr>
<tr>
<td>Total</td>
<td>163</td>
<td></td>
<td>56</td>
<td></td>
</tr>
</tbody>
</table>

Chi-square test $p = 0.2776$

<table>
<thead>
<tr>
<th>Number of C4B genes</th>
<th>Normal Blacks</th>
<th>Black SLE</th>
<th>Chi-square test</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Freq.</td>
<td>N</td>
<td>Freq.</td>
</tr>
<tr>
<td>0</td>
<td>7</td>
<td>0.043</td>
<td>3</td>
<td>0.054</td>
</tr>
<tr>
<td>1</td>
<td>36</td>
<td>0.221</td>
<td>12</td>
<td>0.214</td>
</tr>
<tr>
<td>2</td>
<td>101</td>
<td>0.620</td>
<td>37</td>
<td>0.661</td>
</tr>
<tr>
<td>3</td>
<td>16</td>
<td>0.098</td>
<td>4</td>
<td>0.071</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>0.018</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>Total</td>
<td>163</td>
<td></td>
<td>56</td>
<td></td>
</tr>
</tbody>
</table>

Chi-square test $p = 0.812$

<table>
<thead>
<tr>
<th>RCCX haplotype</th>
<th>Normal Blacks</th>
<th>Black SLE</th>
<th>Chi-square test</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>LL</td>
<td>55</td>
<td>0.169</td>
<td>16</td>
<td>0.145</td>
</tr>
<tr>
<td>LS</td>
<td>207</td>
<td>0.635</td>
<td>67</td>
<td>0.609</td>
</tr>
<tr>
<td>S</td>
<td>15</td>
<td>0.046</td>
<td>9</td>
<td>0.082</td>
</tr>
<tr>
<td>L</td>
<td>21</td>
<td>0.064</td>
<td>11</td>
<td>0.100</td>
</tr>
<tr>
<td>LSL</td>
<td>3</td>
<td>0.009</td>
<td>1</td>
<td>0.009</td>
</tr>
<tr>
<td>LSS</td>
<td>18</td>
<td>0.055</td>
<td>5</td>
<td>0.045</td>
</tr>
<tr>
<td>LLL</td>
<td>3</td>
<td>0.009</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>Others*</td>
<td>4</td>
<td>0.012</td>
<td>1</td>
<td>0.009</td>
</tr>
<tr>
<td>Total</td>
<td>326</td>
<td></td>
<td>110</td>
<td></td>
</tr>
</tbody>
</table>

Chi-square test $p = 0.669$

<table>
<thead>
<tr>
<th>Number of L genes</th>
<th>Normal Blacks</th>
<th>Black SLE</th>
<th>Number of S genes</th>
<th>Normal Blacks</th>
<th>Black SLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>370</td>
<td>0.577</td>
<td>117</td>
<td>0.568</td>
<td>271</td>
<td>0.423</td>
</tr>
<tr>
<td>L gene: S gene</td>
<td>1.365</td>
<td>1.315</td>
<td>C4 gene index</td>
<td>3.957</td>
<td>3.750</td>
</tr>
<tr>
<td>C4A gene index</td>
<td>2.129</td>
<td>2.000</td>
<td>C4B gene index</td>
<td>1.828</td>
<td>1.750</td>
</tr>
<tr>
<td>C4A gene: C4B gene</td>
<td>1.165</td>
<td>1.143</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Others include LSSS, SS and SLS.

Table 4.4 A comparison of the variations in total C4, C4A and C4B gene dosages and RCCX modules between normal African Americans and African American SLE patients.
<table>
<thead>
<tr>
<th></th>
<th>Caucasian SLE</th>
<th></th>
<th>Black SLE</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Freq.</td>
<td>N</td>
<td>Freq.</td>
</tr>
<tr>
<td><strong>Number of C4 genes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>16</td>
<td>0.065</td>
<td>1</td>
<td>0.018</td>
</tr>
<tr>
<td>3</td>
<td>88</td>
<td>0.358</td>
<td>18</td>
<td>0.321</td>
</tr>
<tr>
<td>4</td>
<td>129</td>
<td>0.524</td>
<td>31</td>
<td>0.554</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>0.049</td>
<td>6</td>
<td>0.107</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0.004</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>246</td>
<td></td>
<td>56</td>
<td></td>
</tr>
<tr>
<td><strong>Chi-square test</strong></td>
<td></td>
<td></td>
<td><strong>p</strong> = 0.2995</td>
<td></td>
</tr>
</tbody>
</table>

|                      | Caucasian SLE |                  | Black SLE |                  |
|                      | N   | Freq. | N   | Freq. |
| **Number of C4A genes** |     |       |     |       |
| 0                    | 13  | 0.053 | 0   | 0.000 |
| 1                    | 67  | 0.272 | 9   | 0.161 |
| 2                    | 130 | 0.528 | 40  | 0.714 |
| 3                    | 27  | 0.110 | 5   | 0.089 |
| 4                    | 8   | 0.033 | 2   | 0.036 |
| 5                    | 1   | 0.004 | 0   | 0.000 |
| **Total**            | 246 |       | 56  |       |
| **Chi-square test**  |     |       | **p** = 0.0846 |     |

|                      | Caucasian SLE |                  | Black SLE |                  |
|                      | N   | Freq. | N   | Freq. |
| **Number of C4B genes** |     |       |     |       |
| 0                    | 8   | 0.033 | 3   | 0.054 |
| 1                    | 53  | 0.215 | 12  | 0.214 |
| 2                    | 176 | 0.715 | 37  | 0.661 |
| 3                    | 8   | 0.033 | 4   | 0.071 |
| 4                    | 1   | 0.004 | 0   | 0.000 |
| **Total**            | 246 |       | 56  |       |
| **Chi-square test**  |     |       | **p** = 0.5651 |     |

|                      | Caucasian SLE |                  | Black SLE |                  |
|                      | N   | Freq. | N   | Freq. |
| **RCCX haplotype**   |     |       |     |       |
| LL                   | 219 | 0.445 | 16  | 0.145 |
| LS                   | 123 | 0.250 | 67  | 0.609 |
| S                    | 97  | 0.197 | 9   | 0.082 |
| L                    | 31  | 0.063 | 11  | 0.100 |
| LSL                  | 4   | 0.008 | 1   | 0.009 |
| LSS                  | 7   | 0.014 | 5   | 0.045 |
| LLL                  | 7   | 0.014 | 0   | 0.000 |
| Others(LSSS & SS)    | 4   | 0.008 | 1   | 0.009 |
| **Total**            | 492 |       | 110 |       |
| **Chi-square test**  |     |       | **p** = 5.962E-13 |     |

|                      | Caucasian SLE |                  | Black SLE |                  |
|                      | N   | Freq. | N   | Freq. |
| **Number of L genes** |     |       |     |       |
| 630                  | 0.718 | 117  | 0.568 |
| **Number of S genes** |     |       |     |       |
| 248                  | 0.282 | 89   | 0.432 |
| **L gene: S gene**   | 2.540 | 1.315 |
| **C4 gene index**    | 3.569 | 3.750 |
| **C4A gene index**   | 1.809 | 2.000 |
| **C4B gene index**   | 1.760 | 1.750 |
| **C4A gene: C4B gene** | 1.028 | 1.143 |

Table 4.5 A comparison of the variations in total C4, C4A and C4B gene dosages and RCCX modules between Caucasian and African American SLE patients.
CHAPTER 5

GENERAL DISCUSSION

The immune system is a unique system that functions in the recognition of self from nonself, therefore defends our body against infection. Autoimmune disease is a group of disease caused by the activation of autoreactive T and/or B cells in the absence of discernible cause. SLE is a prototype of systemic autoimmune disease that affects many organs and systems. Most autoimmune diseases including SLE are considered as complex disease, where both environmental factors and genetic factors are involved in disease manifestation. Complement C4, an essential component of complement system that is part of innate immune response, is related to SLE development.

C4 is one of the most polymorphic molecules in the human genome. The frequent variations in gene number, gene size and nucleotide polymorphism of C4 may lead to a wide range of C4 protein levels in the circulation and also in the local tissues, which could in turn affect the immune functions in different individuals. Studies suggested that deficiency of C4 was associated with a variety of autoimmune and infectious diseases. However, the previous “two-locus” model of C4 assumed there were at most two copies of C4 genes, a C4A in the first locus and a C4B in the second locus, on each chromosome 6. This model had been used for over 20 years for the study of C4 in humans. In fact,
three or four copies of functional C4 genes on one C4 haplotype were observed. Through analysis of 150 normal female Caucasians, our laboratory found that the number of C4 genes in a diploid genome varied from 2 to 6. Such C4A-C4B configuration only accounted for 55% in this population. Therefore, an accurate account of the variations of C4 in both normal ethnic populations and race-matched SLE disease populations were in an urgent need.

5.1 Genetic complexity and protein polymorphism of human complement components C4A and C4B in Caucasians

The first goal of my dissertation was to not only validate and extend the results of our previous study with the 150 female Ohioan Caucasians, but also to provide an essential link between the C4 genetic variations with its phenotypic and functional diversities. In a study population of 128 healthy Caucasian subjects from Hungary, the C4 polygenic and gene size variations and their relationships with serum C4A and C4B protein concentrations and hemolytic activities were determined. It was observed that the number of C4 genes in subjects varied from 2 to 5, with 77.4% of the C4 genes being the long form. There were strong positive correlations between total C4, C4A and C4B gene dosages with serum C4A and C4B protein concentrations, respectively. The number of long C4 genes was positively correlated with C4A protein levels, while the number of short C4 genes with C4B protein levels. Higher C4 serum protein levels were detected in subjects with short C4 genes than those with long genes only. Unexpectedly, it was observed that body mass index appeared to affect the C4 serum levels, particularly in the individuals with medium or high C4 gene dosages. Most statistical analyses were performed using unpaired t test. Therefore, C4 gene dosages, gene size, and probably
body mass index, can all affect C4 protein levels in the serum. The positive correlation of long genes with C4A and short genes with C4B proteins indicate that most C4A are long genes and C4B are short genes. Higher C4 serum protein levels in short C4 genes suggest a negative epistatic effect of the retroviral sequence on C4 expression. Intriguingly, the positive impact of short C4 genes on C4 expression seemed more far-reaching because it affected not only C4B, but also C4A protein levels. In subjects with bimodal/bimodal structures including two C4A and two C4B genes, individuals with LS/LS expressed both higher C4B and C4A serum protein levels than those with LL/LL. The exact mechanism how short C4B gene in the second locus affects the expression of the long C4A gene in the first locus is unknown.

C4 hemolytic activity is a functional assay that measures the ability of C4 protein to lyse the antibody-sensitized sheep erythrocytes. It was noticed in the same Hungarian population that there was positive correlation between serum C4 protein concentrations and C4 hemolytic activities, indicating that serum C4A and C4B proteins manifested similar activities in the fluid-phase hemolytic assays. It appears that the differential reactivities between C4A and C4B for bindings to amino group-containing antigens such as immune complex and hydroxyl group-containing antigens such as sheep erythrocytes could be demonstrated only when the C4A and C4B proteins were purified by an immunochemical method or resolved by gel electrophoresis (Awdeh and Alper 1980, Gatenby et al 1990, Gibb et al 1993, Paul et al 1988, Reilly 1999, Reilly and Mold 1997, Schifferli et al 1987). A similar reactivity between C4A and C4B was shown in assay systems when whole human sera were used as a resource of C4A or C4B protein as what was performed in our study, or in reconstituted assays when human C4A- or C4B-
(Christiansen et al 1991, Juan DD et al 1993, Naves et al 1998, Ranford et al 1987, Reveille et al 1995, Reveille et al 1998). Future studies to determine the proteins interacting with C4 and their possible regulatory mechanism would be particularly beneficial to understand the physiological role of C4 in humans. It was observed that the prevalence and severity of autoimmune diseases were different among different ethnic groups. Since C4 is a highly polymorphic molecule and many studies suggest its association with autoimmune diseases, further in-depth genetic, functional and structural characterizations of C4 variations in different ethnic groups may help us understand the possible roles of C4 in disease prevalence, severity and response to therapies.

5.2 Molecular basis of complete complement C4 deficiency in human SLE and kidney disease patients

Complete C4A and C4B deficiencies in human are extremely rare, and only 26 cases had been reported so far. All but one HLA-typed, C4-deficient individual experienced symptoms related to immune complex or clearance disorders such as lupus or lupus-like disease, nephritis or kidney disease. The second goal of my dissertation was to elucidate the molecular basis of complete C4A and C4B deficiencies in these patients. The first patient that I participated in studies was from the US. The patient experienced multiple episodes of infection and he was diagnosed with SLE at age 9. He was homozygous for HLA A2 B12 DR6 with LS RCCX haplotype including a defective long C4A gene and a defective short C4B gene. PCR analysis and DNA sequencing by Kristi L. Rupert in our laboratory revealed that the mutant C4A contained a 2-bp insertion in exon 29, while the C4B mutant had a novel, single C-nucleotide deletion in exon 13 (Rupert et al 2002). From the Alpine region of northern Italy, seven complete C4
deficient patients from four different families were studied and summarized in my dissertation. They were homozygous for either HLA A24 Cw7 B38 DR13 with monomodular L RCCX haplotype including one long C4A gene; or HLA A30 B18 DR7 with bimodular SS RCCX haplotype containing two short C4B genes. Sequencing studies revealed that a novel 2-bp deletion in exon 13 of the long C4A mutant gene, and the identical G to A substitution at 5’ donor site of intron 28 in the two short C4B genes. Altogether, the mutations from 12 reported complete C4-deficient subjects were determined. The mutation analysis suggested two hot spots with deleterious mutations in C4AQ0 and C4BQ0: one is located at exon 13; the other is located within a 2.6 kb genomic region spanning exon 20 to exon 29. Definitive SSP-PCR and RFLPs had been created to facilitate screening of the C4 gene mutations in SLE patients. The results in this study could help epidemiological studies of SLE associated with complement C4A and C4B deficiencies. Besides C4 mutations that lead to no C4 protein production in the circulation, C4 polymorphisms may also affect their functions. One typical example is C4A6 that has a very low C5 convertase activity caused by the lack of C5 binding to the C4b-C3b or to the C4b-C4b complex (Kinoshita et al 1989). Sequencing of the C4A6 gene revealed the R458W substitution on the β-chain (Anderson et al 1992) that was confirmed to confer low C5 convertase activity by site-directed mutagenesis (Ebanks et al 1992). It would be informative to elucidate the structural basis of all polymorphic variants among different ethnic groups, and to investigate the potential functional variations of these protein isoforms interacting with a variety of targets, or receptors, which include the C1 complex, C3, C5, the regulatory proteins C4bp, MCP, DAF and CR1, the receptors CR1, C-reactive protein and the serum amyloid protein. It would help
further therapeutic intervention of C4-related diseases using nonimmunogenic C4 to reduce the possibility of inducing adaptive immune response.

5.3 The roles of complement C4 gene dosage variation and C4A deficiency in SLE

While a complete C4 deficiency almost always leads to lupus or lupus-like diseases, it has been controversial about the role of C4 quantitative variations in SLE disease association. The third goal of my dissertation was to accurately determine C4 polygenic variation and the association of partial C4A deficiency with SLE. There were conflicting results supporting or refuting the association of partial C4A deficiency with SLE. Two potential problems greatly hindered the study of C4 in SLE in the past. First, most previous studies were based on phenotypic analysis using the “two-locus” model, where the unequal band intensities of C4A and C4B proteins in allotyping gels were assigned as partial C4A or C4B deficiencies. However, with C4 gene dosage varying from 2 to 6 in human populations, unequal C4A and C4B gene dosages, or unequal expression levels among C4A and C4B genes would lead to an apparent C4A or C4B “partial deficiency”. Therefore, an accurate analysis of C4 polygenic variations in SLE can help unambiguously assess the role of total C4 and C4A deficiencies in susceptibility to lupus. Second, C4 is located in the gene dense MHC region, where susceptibility to many autoimmune diseases including SLE is associated with the presence of particular haplotypes of genes. The strong linkage disequilibrium makes it extremely difficult to segregate the contribution of an individual gene to disease manifestation. In this dissertation, 209 female Caucasian SLE patients, 56 African American SLE patients and their unaffected family members were studied. It was shown that deficiencies of total C4 and C4A genes were increased in both Caucasian and African American SLE, although
the statistical analysis in African Americans was not significant due to the limited sample size. The C4A gene mutation rate increased from 2% in unrelated controls to 5.79% in SLE patients. Intriguingly, protein levels of total C4, C4A and C4B decreased in SLE patients, even within the same gene dosage group. The C4 consumption, indicated by C4 split product C4a, also increased in SLE patients. Therefore, the low protein levels of total C4, C4A, and to a less extent, C4B in SLE were caused by deficiencies of total C4 and C4A genes, C4A mutations, and low C4 expression or high C4 consumption. The C4 and C4A deficiencies in the two different ethnic groups and lower C4 and C4A protein levels caused by different mechanisms underlie the role of C4, especially C4A, in SLE disease manifestation. Mutation studies of two other susceptibility genes C2 and TNFA in MHC class III region suggested they were only secondary risk factors in SLE development.

A particular haplotype, HLA A1 Cw7 B8 DR3, i.e. mono-S RCCX haplotype, is always linked to C4A deficiency. Previous studies suggested this haplotype was associated with many autoimmune diseases including SLE (Candore et al 2002). In our study, this haplotype was increased from 11% in controls to 19.1% in Caucasian SLE. Its frequency in African Americans also increased from 4.6% in controls to 8.2% in patients, although the HLA haplotype in African Americans is different. Although this haplotype itself was not sufficient to explain the lower C4A protein levels in SLE patients, its increase in patient populations suggested the special combination of genetic polymorphisms in this haplotype may contribute to SLE disease etiology. In other words, changes in other genes in this specific MHC haplotype may be involved in lupus development. Using microarray chips to study the gene expression profile in the MHC
region from individuals with homozygous mono-S RCCX haplotypes may be informative to understand the possible involvement of other MHC genes in lupus. More fundamentally, a customized microarray chip that includes genes in MHC, complement genes, Fc receptor genes, apoptotic factor genes, lymphokines, and other genes or loci linked to lupus, could be designed to analyze the SLE patients, unaffected family members and unrelated controls. The results would yield relevant information of possible genetic defects in lupus, and possibly help predicting disease progression. Functional studies of susceptibility genes using appropriate animal models could allow an assessment of their roles in the disease processes.

For the association of C4 with SLE, several future studies would be necessary. First, a suitable model system is important to study the role of C4 in SLE development. Instead of using mice as models, primates that have polygenic and gene size variations of C4 would be more appropriate as it better mimics the C4 genetic diversity in humans. The interaction of C4 with other molecules in the circulation or on the surface of immune cells, which may be from the lymphoid organs such as bone marrow, spleen and thymus and from lupus targeted organs including kidney and brain, would be of interest. Second, the over-expression of C4 in the local tissue may aggravate the local inflammation and tissue injury in the lupus. Using the cell lines from different tissues and organs, the C4 expressions with and without stimulations could be intensively studied. Understanding the regulation mechanisms of C4 expression in local tissues may enable a manipulation of C4 in different disease stages. Down-regulation of C4 may help reduce tissue injury, while up-regulation of C4 may enhance immunoclearance and deletion of autoreactive B cells. Third, the non-immunogenic C4 fusion protein functioning in the clearance of
immune complexes could be designed using biochemical engineering. If C4A protein was
defective in lupus patients, an infusion of non-immunogenic and functional C4 protein
could be beneficial to reverse the disease progression.

In conclusion, results of this dissertation have firmly established the relationship
of protein levels of total C4, C4A and C4B with their corresponding gene dosages, gene
size, race, and possibly, body mass index of the subject. The determination of molecular
basis of complete C4A and C4B deficiencies contribute to the epidemiological studies of
C4 with C4-related diseases. Our molecular genetic studies of Ohio lupus revealed that
that low gene dosages or protein levels of complement C4 and C4A are primary risk
factors for disease manifestation of SLE.
Bibliography


tenascin X gene, partial cds, cytochrome P450 21-hydroxylase (CYP21B), complement C4 (C4B), G11, helicase (SKI2W), RD, complement factor B (Bf), and complement component C2 (C2) genes, complete cds. GenBank, Accession no. AF019413.


Traustadottir, K. H., K. Steinsson, and K. Erlandsson. 1998. C4AQ0 superimposed on a primary defect increases the susceptibility to systemic lupus erythematosus (SLE) in a family with association between C4AQ0 and SLE. *J. Rheumatol.* 25:2118-2125.


