Sex, MHC and complement C4 in autoimmune diseases

C. Yung Yu¹,²,³,* and Caroline C. Whitacre³

¹Columbus Children’s Research Institute, The Ohio State University, Columbus OH 43205, USA
²Department of Pediatrics, The Ohio State University, Columbus OH 43221, USA
³Department of Molecular Virology, Immunology and Medical Genetics, The Ohio State University, Columbus OH 43205, USA

Autoimmune diseases are estimated to affect 10–50 million people in the United States, and untold millions worldwide. Nearly 80% of all people with autoimmune diseases are women, and a strong association of these diseases with MHC genes has been known for some time. However, very little is known about what causes autoimmune diseases or the factors that lead to disease recurrence. The sex-associated differences in multiple sclerosis (MS) and the mouse model of MS, experimental autoimmune encephalomyelitis (EAE), are associated with MHC genetic background, sex hormone levels and cytokine production. The implication of these factors has aided the identification of new autosomal genetic susceptibility loci. Complete deficiencies of early complement components are strongly associated with systemic lupus erythematosus (SLE) but the role of complement proteins in SLE is not yet clear. Recent data suggest that quantitative and qualitative diversities of the MHC-linked complement C4 among different ethnic groups can be important in the susceptibility and disease severity of SLE.

Sex, race and MHC are important factors associated with human autoimmune diseases. The sex and racial differences in the prevalence of autoimmune diseases are striking. In the US, the incidence of systemic lupus erythematosus (SLE) in women is nine times greater than in men [1], and is three times more frequent in ethnic groups can be important in the susceptibility to autoimmunity. The sex-associated differences in multiple sclerosis (MS) and the mouse model of MS, experimental autoimmune encephalomyelitis (EAE), are associated with MHC genetic background, sex hormone levels and cytokine production. The implication of these factors has aided the identification of new autosomal genetic susceptibility loci. Complete deficiencies of early complement components are strongly associated with SLE but the role of complement proteins in SLE is not yet clear. Recent data suggest that quantitative and qualitative diversities of the MHC-linked complement C4 among different ethnic groups can be important in the susceptibility and disease severity of SLE.

Sexual dimorphism in MS and EAE

In MS, the course of disease and the response to the most common drug in current use, interferon-β (IFN-β), varies between males and females. Women with MS tend to have a remitting/relapsing disease course, whereas men tend to exhibit a primary progressive course. Consequently, women respond more favorably to IFN-β therapy, which is of greater benefit in remitting/relapsing MS [5,6]. The most striking effect of sex on autoimmune disease is during pregnancy, as diseases such as MS and RA show profound improvement during pregnancy, with flares of disease following parturition. By contrast, SLE is reported to worsen during pregnancy [6]. The underlying mechanisms of sex effects in human autoimmune disease are not known at present, but could involve the following: effects mediated directly by genes located on the sex chromosomes; sex hormones acting at the level of the immune system, the target organ or hormone response elements present in a wide variety of genes; hormones of the hypothalamic-pituitary-adrenal (HPA) axis that have sexually dimorphic effects; or a combination of these effects. The greatest focus of attention to date has been on the effects of sex hormones on the development of autoimmune diseases.

Genetics of sex differences in EAE

Defining the immunogenetics of sex differences in autoimmune disease is in the early stages, but some advances have been made, particularly with the mouse model for MS, EAE. A recent survey of murine strains for sex differences in EAE has shown a major effect of MHC (Table 1), as H-2α mouse strains (SJL and ASW) and the autoimmune-prone NZW (H-2b) strain show a clear female predilection for disease; H-2β mouse strains (B10.PL and PL/J) show a male predilection; and others show no sex differences (C57/BL/6 [H-2b] and NOD [H-2k]) [7]. Elegant studies identifying the genetic basis for differing disease subtypes and sexual dimorphism in EAE have been conducted by Teuscher and Blankenhorn [8] using an F2 population derived from the EAE-susceptible SJL/J and EAE-resistant B10.S/DvTe inbred lines. Immunization of a large group of F2 mice with spinal cord homogenate for EAE induction resulted in the definition of five different clinical subtypes of EAE: benign, acute progressive, chronic/nonremitting, remitting/relapsing and monophasic remitting/nonrelapsing. The frequency
traits (QTs) associated with EAE [7]. Ovariectomy shifted differences in clinical and histopathological quantitative the role of gonadal hormones in directing the sex backcross populations [8,9]. Of note is the distribution of with the sexually dimorphic aspects of EAE in the F2 or specific effects. Table 2 shows the genetic loci associated controlling susceptibility to each disease subtype and sex-

Genome exclusion mapping was used to identify the loci associations with nearby loci of immunological QTs for both brain and spinal cord. By contrast, to an acute progressive form, and reduced the histopatho-

lesion severity and demyelination. Interestingly, gona-

crastated males have increased disease severity

of mice showing clinical and histopathological manifesta-

tions of EAE was significantly greater for females than males in this group, with females biased toward a remitting/relapsing disease course and males biased toward a monophasic remitting/nonrelapsing course [8]. Genome exclusion mapping was used to identify the loci controlling susceptibility to each disease subtype and sex-
specific effects. Table 2 shows the genetic loci associated with the sexually dimorphic aspects of EAE in the F2 or backcross populations [8,9]. Of note is the distribution of loci across several autosomal chromosomes and the associations with nearby loci of immunological significance.

An interaction between genetic background and sex hormones was recently observed in the castration of SJL males, which caused an increase in severity of EAE, whereas castration of C57BL/6 males produced no change in disease [10]. The F2 group was used to further examine the role of gonadal hormones in directing the sex differences in clinical and histopathological quantitative traits (QTs) associated with EAE [7]. Ovariectomy shifted the EAE clinical course from remitting/relapsing disease to an acute progressive form, and reduced the histopatho-

dimorphism of histopathological changes observed in the spinal cord [11].

**Genetics of sex differences in Theiler’s virus infection**

Susceptibility to Theiler’s murine encephalomyelitis virus-induced demyelination (TMEVD), another mouse model for MS, has also been studied for sex-specific effects [12]. Composite interval mapping was used to identify quantitative trait loci (QTL) in male and female backcross populations derived from breeding susceptible DBA/2J with resistant BALBc/ByJ mice. A significantly greater proportion of males showed gait abnormalities than females. QTL on chromosomes 1, 5, 15 and 16 were identified for males, whereas two QTL were found for female mice, both located on chromosome 1. The same interval of chromosome 1 contains QTL with opposite effects in male and female populations. The QTL identified on chromosome 16 in males colocalizes with eae11, which is responsible for controlling lesion severity and suscepti-

**Sex differences in cytokine production in MS**

Obvious candidates for exploration of the genetic basis for sex differences in autoimmune disease are cytokine genes. In MS, females secrete higher amounts of IFN-γ in response to myelin proteolipid protein (PLP) than males [13]. Furthermore, females with MS show no IL-5 response to PLP in contrast to males with MS. The IFN-γ gene contains a microsatellite polymorphism in the first intron, and together with two single nucleotide polymorphisms

<table>
<thead>
<tr>
<th>Mouse strains</th>
<th>H-2 haplotype</th>
<th>Distinct features in EAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SJL</td>
<td>s</td>
<td>Female predilection</td>
</tr>
<tr>
<td>ASW</td>
<td>s</td>
<td>Castrated males have increased disease severity</td>
</tr>
<tr>
<td>NZW</td>
<td>d</td>
<td>Female predilection</td>
</tr>
<tr>
<td>B10.PL</td>
<td>u</td>
<td>Male predilection</td>
</tr>
<tr>
<td>PUJ</td>
<td>u</td>
<td>Castrated females have increased disease severity</td>
</tr>
<tr>
<td>C57/BL/6</td>
<td>b</td>
<td>No sex differences</td>
</tr>
<tr>
<td>NOD</td>
<td>g7</td>
<td>Castrated males have no increase in disease severity</td>
</tr>
</tbody>
</table>

**F2 hybrids**

- SJL/J (EAE-susceptible) and B10.S/OvTe (EAE-resistant) Five clinical subtypes
- DBA (TMEVD-susceptible), and Balb/c/2J (TMEVD-resistant) More females than males with clinical and histopathological manifestations

<table>
<thead>
<tr>
<th>Table 1. Sex and MHC in EAE disease manifestations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Locus</strong></td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>eae1</td>
</tr>
<tr>
<td>eae1f1</td>
</tr>
<tr>
<td>eae12</td>
</tr>
<tr>
<td>eae13</td>
</tr>
<tr>
<td>eae17</td>
</tr>
<tr>
<td>eae18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2. Summary of genetic loci shown to be involved in sex differences in EAE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Locus</strong></td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>eae1</td>
</tr>
<tr>
<td>eae1f1</td>
</tr>
<tr>
<td>eae12</td>
</tr>
<tr>
<td>eae13</td>
</tr>
<tr>
<td>eae17</td>
</tr>
<tr>
<td>eae18</td>
</tr>
</tbody>
</table>

*Abbreviations: AIRE, autoimmune regulator; EAE, experimental autoimmune encephalomyelitis; IGF, insulin-like growth factor; MBP, myelin basic protein; MIF, macrophage migration inhibitory factor; SLE, systemic lupus erythematosus; TGF-β1, transforming growth factor-β1; TMEV, Theiler’s murine encephalomyelitis virus-induced demyelination.*
(SNPs) forms two major haplotypes, I and II. Haplotype I is associated with high production of IFN-γ, whereas haplotype II is associated with low production of this cytokine. Interestingly, a series of reports have identified haplotype I as protective in males with MS [14]. Although these results appear to be contradictory, the role of IFN-γ in MS and EAE is both protective and detrimental depending upon the stage of disease [15]. The status of both human and other animal genetic approaches related to MS have been reviewed recently [16].

**MHC and complement in human SLE**

In about half of the autoimmune patient populations, the underlying genetic factors are specific combinations of alleles, genetic markers or haplotypes of the MHC. The MHC was initially defined for its role in self/non-self recognition, graft rejection and control of the immune response. The concept of the human MHC has now evolved to include a physical region spanning ~4700 kb of more than 120 functional genes on chromosome 6 [17]. Genes of the MHC encode not only the well-known class I and class II proteins involved in antigen presentation, but also complement proteins that constitute the C3 convertases (C4a, C4b, C2 and factor B), tumor necrosis factor (TNF)-α and its related proteins lymphotoxin-α and -β, and heat shock protein 70 (HSP70). Many genes with heterogeneous functions have been mapped to the extraordinarily dense segment between the class I and class II regions [18,19].

**Two genomic regions in the MHC with variable gene compositions**

In addition to the extremely polymorphic sequences for some MHC class I and class II genes that can vary 1–10% among individuals [20,21], another remarkable inherent genetic feature of the human MHC is the presence of two genomic regions that vary in length and gene content. In the class II region between DRB1 and DRB9, five polymorphic length variants with 0 (DR8), 1 (DR1), 2 (DR51 and DR52) or 3 (DR53) functional or non-functional genes for the beta chain of DR (DB) can be present [22]. In the class III region, modular duplications involving four consecutive genes are present. Each module contains four contiguous genes coding for nuclear kinase RP [also known as serine/threonine protein kinase (Stk)19], complement C4, cytochrome P450 21-hydroxylase (CYP21) and the extracellular matrix protein tenascin-X (TNX), collectively forming the RCCX module (RCCX has been defined here by underlining the genes involved) [23,24]. The duplicate module contains two intact genes, CYP21 and C4; and two incomplete gene fragments, RP2 and TNXA (Figure 1). The CYP21 in each duplicated segment is either a non-functional CYP21A pseudogene or a functional CYP21B. Each duplicated C4 gene can code for an acidic C4A or a basic C4B protein. Moreover, the C4 gene is either a long or a short gene. This C4 gene size dichotomy is due to the integration of the human endogenous retrovirus into the ninth intron of long C4 genes [HERV-K(C4)] [25]. The configuration of HERV-K(C4) is opposite to that of C4, implying that RNA molecules complementary to retroviral sequences are produced whenever a long C4 gene is transcribed.

**MHC-linked complement C4 genes in human SLE**

Research into the role of the MHC in autoimmune diseases has focused mainly on gene polymorphisms, and peptide binding and antigen presentation by class I and class II proteins. Proteins encoded by genes in the class III region, including the MHC-linked complement proteins, might also be involved in the pathogenesis of autoimmune disease.

Complete C4A and C4B deficiencies are extremely rare in human populations, and only 26 subjects have been documented worldwide. Of these, 14 were diagnosed with SLE (according to the American College of Rheumatology criteria) and 11 had glomerulonephritis or lupus-like symptoms, such as photosensitivity and skin lesions [26,27]. This strong association shows that deficiency of complement C4 is likely to be an etiological factor in SLE. In fact, low serum concentrations of complement C4 are frequently observed in SLE patients. Many studies over the past twenty years have revealed that homozygous or partial deficiency of C4A (characterized by lower expression of C4A relative to C4B in plasma or serum) is present in 32–55% of SLE patients. Such values are two to five times greater than the matched control groups [27]. Intriguingly, Spanish, Mexican and Australian Aborigine SLE patients have a high frequency of C4B deficiency (instead of C4A deficiency), or lower C4B plasma levels relative to C4A [28,29]. Thus, it appears that there is a difference in the requirement of C4A or C4B proteins in the autoimmunity process among individuals from various races. How the differential expression of C4A and C4B proteins contribute to the pathogenesis of SLE deserves in-depth studies.

Complement C4 and its related protein C3 are remarkable for their ability to form covalent bonds with nearby target surfaces, including microbial cell membranes, viral capsids and host cell membranes [30]. After activation of C4, a cascade of reactions leads to the formation of the membrane attack complex and production of biologically active peptides including potent anaphylatoxins. The deposition of activated C4 molecules on microbial or other surfaces facilitates the clearance of immune complexes by red blood cells and macrophages (through binding to complement receptor CR1 and/or Fc receptors). C4A and C4B differ in their chemical reactivities towards antigens; activated C4A prefers to form an amide bond with protein antigens, such as immune complexes, whereas activated C4B reacts rapidly with a hydroxyl group to form an ester linkage with carbohydrate antigens, such as bacterial cell walls [31,32]. It is therefore postulated that the longer half-life and preferred binding to protein targets for activated C4A are relevant features for the engagement of complement C4 in the secondary immune response and in tolerance [33].

**Great racial difference in MHC haplotypes with long and short C4 genes and RCCX modular variants**

In the MHC, specific combinations of polymorphic sequences spanning thousands of kilobases, together with polygenic variations of DRB and RCCX, tend to be segregated and inherited together as a block with a very
low frequency of productive recombinations [34]. Such a phenomenon gives rise to the MHC ancestral haplotypes, which are usually specific among different human races and ethnic groups. We have studied over 2000 healthy subjects in total, including SLE patients and family members from diverse races and ethnic groups, for the composition of their RCCX modules and C4 genes. We found that between one and four copies of C4 genes can be present in the MHC. Between two and seven copies of C4 genes are present in a diploid genome. Fourteen variants of RCCX modules in the MHC with different combinations of long and short C4 genes have been found [32,35–38] (Figures 1 and 2). In Caucasians, 76% of C4 genes exhibit the long form and 24% have the short form. In African Americans, the frequency of short C4 genes increases to 42%. Bimodular long-long (LL) is the most common RCCX structure in Caucasians, whereas bimodular long-short (LS) is the most prevalent structure in Blacks and Orientals. Monomodular short (S) without a C4A gene has a frequency of ~1% in Caucasians, but only ~1% in Asian populations (Figure 2). Approximately 20% of healthy Caucasians have a heterozygous deficiency of C4A, and 20–30% have a haploinsufficiency of C4B. At the other end of the spectrum, 12–18% of Caucasians have five or six C4 genes and high levels of C4A and/or C4B proteins [35,36]. Chinese and Asian Indians have relatively higher C4 gene dosages than Caucasians, as 24–31% of these ethnic groups have five to seven C4 genes in a diploid genome.

Quantitative diversities of C4A and C4B proteins

The plasma or serum C4 protein levels in healthy individuals vary widely from 15 to 80 mg/dL. In most SLE patients, C4 levels fluctuate with time, and at times can be less than 10 mg/dL. In healthy populations, it appears that C4 gene dosage, gene size and the body mass index of an individual contribute to the quantitative variation of plasma C4 protein levels. Remarkably, the plasma protein concentrations in individuals with four long C4 genes (LL/LL) are consistently ~40% lower than in those with two long and two short C4 genes (LS/LS), suggesting that HERV-K(C4) downregulates the expression of C4 proteins from long C4 genes [36]. Indeed, it was shown that cloned sequences from the 3’0 long terminal repeat of the HERV-K(C4) have promoter activity and direct synthesis of antisense transcripts complementary to that of human C4 [25,39]. Consistent with the high frequency of short C4 genes in Blacks, the mean C4B serum concentrations in Blacks are ~40% higher than that in Whites [40]. Although homozygous or heterozygous deficiencies of C4A appear to be a disease susceptibility factor of SLE in many races, high expression levels of C4B could be a disease severity factor. This is because high C4B protein levels might aggravate the complement-mediated tissue injuries during disease flares. We are now studying the relationship between gene dosage, gene size variations and C4 protein levels in Black and Asian SLE patients. It is anticipated that the results will shed light on why the prevalence and severity
of SLE in Blacks and Asians are both higher than in Whites.

The correlation of C4 plasma or serum protein levels with body mass index might reflect the biosynthesis of complement proteins including C4 by adipose tissues. Recently, it has been shown that serum C4 levels in females correlate not only with omental adipose tissue mass, as in the males, but also with subcutaneous adipose tissue and total fat depot, wherein males show no correlation [41]. It is of interest to determine if such differential expression patterns of complement C4 and other effector molecules of immune response in subcutaneous and other adipose tissues play a role in the pathogenesis of female-dominant autoimmune diseases.

**MHC ancestral haplotypes and autoimmune diseases**

In Caucasians, two MHC ancestral haplotypes are frequently associated with autoimmune diseases. The AH8.1 haplotype has a monomodular short RCCX present between HLA A*01–Cw*07–B*08 in the class I region, and DRB1*0301–DRB3*0101–DQA1*0501–DQB1*0201–DPA1*01–DPB1*0301 in the class II region. This haplotype has no HERV-K(C4) in the C4B gene and the C4A gene is absent. This haplotype is strongly associated with SLE and type 1 diabetes [34]. Complete genomic DNA sequences for these two haplotypes have recently been published, and sequences for six other MHC haplotypes are being mapped and sequenced (The MHC haplotype project; http://www.sanger.ac.uk/HGP/Chr6/MHC/index.shtml) [42]. Further in-depth characterization of MHC haplotypes among different races and diversities of protein functions including those from the class III region are crucial for our understanding of the MHC in autoimmunity.

**Conclusions**

Recent advances in genome sequences, SNP and haplotype mapping, and genome wide analyses of tissue-specific gene expression patterns provide a solid framework that promises to substantially facilitate the identification and characterization of candidate genes involved in the etiology of complex diseases. In this review, we have focused on two specific examples, MS/EAE and SLE, where genetic studies have focused on elucidating sex differences and disease susceptibility, respectively. The greatest challenge now is how to meticulously investigate the physiological functions and genetics of the candidate genes and to establish the cause-effect relationships in the multifactorial pathogenesis process. It appears to us that working towards an understanding of how sex, MHC and complement C4 are involved in so many autoimmune diseases prevalent in women of reproductive ages provides an ideal model system to decipher...
the many components involved in autoimmune diseases. This also gives us the opportunity to elucidate how quantitative and qualitative variations of immune effector proteins, in response to stimuli by sex hormones and microbial infections, contribute to differences in disease susceptibility, disease severity and response to therapies.

Acknowledgements

We would like to express our gratitude to our colleagues for collaborations and for providing experimental data described in this manuscript. This work is supported by National Institutes of Health grants SRO1 AR050078 from the National Institute of Arthritis, Musculoskeletal and Skin Diseases (CYY), 1P01 DK55546 from National Institute of Digestive, Diabetes and Kidney Diseases (CYY/Hebert LA), and R01 NS48316 from National Institute of Neurological Disorders and Stroke (CCW), and National Multiple Sclerosis Society grant RG3091 (CCW).

References

4 Onkamo, P. et al. (1999) Worldwide increase in incidence of Type I diabetes—the analysis of the data on published incidence trends. Diabetologia 42, 1395–1403
12 Butterfield, R.J. et al. (2003) Sex-specific quantitative trait loci govern susceptibility to Theiler’s murine encephalomyelitis virus-induced demyelination. Genetics 163, 1041–1046
21 Horton, R. et al. (1998) Large-scale sequence comparisons reveal unusually high levels of variation in the HLA-DQB1 locus in the class II region of the human MHC. J. Mol. Biol. 282, 71–97
33 Fino, O. et al. (1992) Structural differences between the two human complement C4 isotypes affect the humoral immune response. J. Exp. Med. 175, 537–543