373e The Major Histocompatibility Complex

Gerald T. Nepom

THE HLA COMPLEX AND ITS PRODUCTS

The human major histocompatibility complex (MHC), commonly called the human leukocyte antigen (HLA) complex, is a 4-megabase (Mb) region on chromosome 6 (6p21.3) that is densely packed with expressed genes. The best known of these genes are the HLA class I and class II genes, whose products are critical for immunologic specificity and transplantation histocompatibility, and they play a major role in susceptibility to a number of autoimmune diseases. Many other genes in the HLA region are also essential to the innate and antigen-specific functioning of the immune system. The HLA region shows extensive conservation with the MHC of other mammals in terms of genomic organization, gene sequence, and protein structure and function.

The *HLA class I genes* are located in a 2-Mb stretch of DNA at the telomeric end of the HLA region (Fig. 373e-1). The classic (MHC class Ia) HLA-A, -B, and -C loci, the products of which are integral participants in the immune response to intracellular infections, tumors, and allografts, are expressed in all nucleated cells and are highly polymorphic in the population. *Polymorphism* refers to a high degree of allelic variation within a genetic locus that leads to extensive variation between different individuals expressing different alleles. More than 2000 alleles at HLA-A, nearly 3000 alleles at HLA-B, and more than 1700 at HLA-C have been identified in different human populations, making this the most highly polymorphic segment known within the human genome. Each of the alleles at these loci encodes a *heavy chain* (also called an *α chain*) that associates noncovalently with the nonpolymorphic light chain β_2*-microglobulin*, encoded on chromosome 15.

The nomenclature of HLA genes and their products is based on a revised World Health Organization (WHO) nomenclature, in which alleles are given a single designation that indicates locus, allotype, and sequence-based subtype. For example, *HLA-A*02:01* indicates subtype 1 of a group of alleles that encode HLA-A2 molecules. Subtypes that differ from each other at the nucleotide but not the amino acid sequence level are designated by an extra numeral (e.g., *HLA-B*07:02:01* and *HLA-B*07:02:02* are two variants of *HLA-B*07:02*, both encoding the same HLA-B7 molecule). The nomenclature of class II genes, discussed below, is made more complicated by the fact that both chains of a class II molecule are encoded by closely linked HLA-encoded loci, each of which may be polymorphic, and by the presence of differing numbers of isotypic DRB loci in different individuals. It has become clear that accurate HLA genotyping requires DNA sequence analysis, and the identification of alleles at the DNA sequence level has contributed greatly to the understanding of the role of HLA molecules as peptide-binding ligands, to the analysis of associations of HLA alleles with certain diseases, to the study of the population genetics of HLA, and to a clearer understanding of the contribution of HLA differences to allograft rejection and graft-versus-host disease. Current databases of HLA class I and class II sequences can be accessed by the Internet (e.g., from the IMGT/HLA Database, *http://www.ebi.ac.uk/imgt/hla*), and frequent updates of HLA gene lists are published in several journals.

The biologic significance of this MHC genetic diversity, resulting in extreme variation in the human population, is evident from the perspective of the structure of MHC molecules. As shown in Fig. 373e-2, the MHC class I and class II genes encode MHC molecules that bind small peptides, and together this complex (pMHC; peptide-MHC) forms the ligand for recognition by T lymphocytes, through the antigen-specific T cell receptor (TCR). There is a direct link between the genetic variation and this structural interaction: The allelic changes in genetic sequence result in diversification of the peptide-binding capabilities of each MHC molecule and in differences for specific TCR binding. Thus, different pMHC complexes bind different antigens and are targets for recognition by different T cells.

The class I MHC and class II MHC structures, shown in Fig. 373e-2 *B*, *C*, are structurally closely related; however, there are a few key differences. While both bind peptides and present them to T cells, the binding pockets have different shapes, which influence the types of immune responses that result (discussed below). In addition, there are structural contact sites for T cell molecules known as CD8 and CD4, expressed on the class I or class II membrane-proximal domains, respectively. This ensures that when peptide antigens are presented by class I molecules, the responding T cells are predominantly of the CD8 class, and similarly, that T cells responding to class II pMHC complexes are predominantly CD4.

The nonclassic, or class Ib, MHC molecules, HLA-E, -F, and -G, are much less polymorphic than MHC Ia and appear to have distinct functions. The HLA-E molecule has a peptide repertoire displaying signal peptides cleaved from classic MHC class I molecules and is the major self-recognition target for the natural killer (NK) cell–inhibitory receptors NKG2A or NKG2C paired with CD94 (see below and Chap. 372e). This appears to be a function of immune surveillance, because loss of MHC class I signal peptides serves as a surrogate marker for injured or infected cells, leading to release of the inhibitory signal and subsequent activation of NK cells. HLA-E can also bind and present peptides to CD8 T cells, albeit with a limited scope, as only three HLA-E alleles are known. HLA-G is expressed mainly in stem cells and in extravillous trophoblasts, the fetal cell population directly in contact with maternal tissues. It binds a wide array of peptides, is expressed in six different alternatively spliced forms, and provides inhibitory signals to both NK cells and T cells, presumably in the service of maintaining maternofetal tolerance. Pathologic expression in cancer and infections may also deliver a similar inhibitory immunologic function; 16 HLA-G alleles have been identified. The protein product of HLA-F is found mainly intracellularly, and the function of this locus, which encodes four alleles but has multiple transcriptional variations, remains largely unknown.

Additional class I–like genes have been identified, some HLA-linked and some encoded on other chromosomes, that show only distant homology to the class Ia and Ib molecules but share the three-dimensional class I structure.

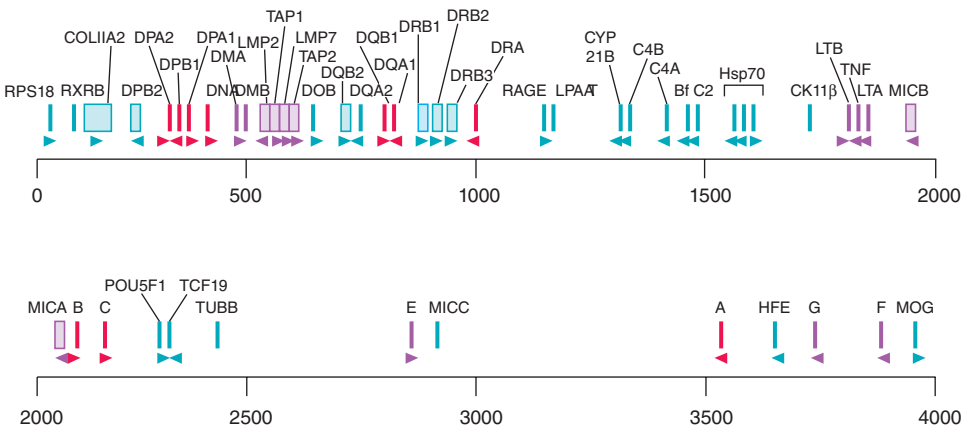

FIGURE 373e-1 Physical map of the HLA region, showing the class I and class II loci, other immunologically important loci, and a sampling of other genes mapped to this region. Gene orientation is indicated by *arrowheads*. Scale is in kilobase (kb). The approximate genetic distance from DP to A is 3.2 cM. This includes 0.8 cM between A and B (including 0.2 cM between C and B), 0.4–0.8 cM between B and DR-DQ, and 1.6–2.0 cM between DR-DQ and DP.