Immunology – Dr. John J. Haddad

Chapter 7 – Major Histocompatibility Complex

The major histocompatibility complex (MHC) is a group of genes that governs tumor and tissue transplantation between individuals of a species.

It was discovered in mice, but all mammalian species have a MHC.

In humans, the complex if genes is called HLA (for human leukocyte antigens). In mice, it is called H-2 (for the group II cell surface antigens discovered by Gorer). Many genes of the MHC encode cell surface membrane-spanning proteins.

Approximately 0.5% recombination can be shown to occur within the extended locus encoding the MHC genes. As we shall see, there are “hot spots” for recombination.

Much of what we know about the MHC of mice was learned by a combination of mouse genetics and the use of complex antisera made by immunizing one individual with cells from another. These are called alloantisera.

You might consider the relationship of such antisera to the anti-allotypes we talked about in discussing immunoglobulins.

Now we know the nucleotide sequence of all genes encoded in the MHC of mice, and have identified genes whose existence was not even suspected from the genetic and immunological approaches.

Inbred mouse strains

Inbred strains are strains of mice within which every mouse has an identical genetic makeup. They were derived beginning around 1900 by crossing pairs of mice and then brother-sister mating their progeny, and then brother-sister mating the next generation, etc. Many lines died out in doing this, probably because they became homozygous for defective alleles at one or more loci. However, some survived to become inbred strains. (Table 7-1).

It started out as a hobby among so-called mouse fanciers. Then people realized they would be great for medical research. It was shown early on that you could exchange skin and other cells from mouse to mouse within an inbred strain successfully, but generally not between different inbred strains.

Genetic, alloantisera and skin transplantation studies showed that the mouse MHC could be divided roughly into five regions.

$$K \quad IA \quad IE \quad S \quad D$$

Importantly, these MHC genes were found to be highly polymorphic --- that is, many alleles could be defined at the genes characterizing each of these major regions. Since only 0.5% genetic recombination was observed between the K and D regions, the alleles at the five regions are passed on to progeny together and they define a haplotype.

In Table 7-1, for example, the CBA strain (as well as AKR, C3H, B10BR and C57BR strains) has the H-2$^k$ (or $k$) haplotype containing the $K^k$, $IA^k$, $IE^k$, $S^k$ and $D^k$ alleles. $K^k$ is spoken of as “K of $k$” and $IA^k$ is spoken of as “IA of $k$.” You should become fluent in this terminology.

Figure 7-1 shows a simplified but useful view of the MHC regions of mouse and humans and a few of the genes encoded in each region.

Congenic strains of mice

You can produce strains of mice which are genetically identical to each other but which differ at a single locus or a set of very closely linked loci (for instance, the MHC). Such strains are called congeneric strains. You can produce
congenic strains of mice differing only at the MHC by the cross-intercross method shown in Figure 7-2/3. In that figure, “the B haplotype is being crossed onto the A genetic background.” You can accomplish the same goal by a simple backcross approach if you have the appropriate antiserum to test progeny for the B haplotype.

Crossing of two mice congenic at the MHC (see B10 and B10.A in Figure 7-4) can give rise to recombinant congenic mice that can be used to establish strains. Experiments done with mice differing only at one or more of the regions within the MHC have revealed much about the function on proteins encoded by each region.

Note: Although similar genetic studies cannot be carried out in humans, evidence obtained in many ways has shown that the human MHC, while organized somewhat differently, is also highly polymorphic and also encodes proteins homologous to and with similar functions to those of mice.

Now we’ll talk about the molecules encoded by the MHC.

**Linkage disequilibrium**

If alleles at different loci in the MHC are inherited according to their frequency in the population, you would expect that if an allele at one locus (e.g., HLA-A1) occurs at a frequency of 16% in the population, and an allele at another (e.g., HLA-B8) is a 9%, then the incidence of inheriting both alleles would by 0.16 x 0.09 = 1.4%. However, the actual incidence observed is much higher at 8%. This is what is meant by linkage disequilibrium What would cause this?

1. Not enough generations have occurred to allow recombination to establish equilibrium among the alleles present in the founders of the population.
2. Recombination to randomize alleles may fail to occur because the nucleotide sequences separating them may resist recombination or because the genes are so close together.
3. Certain combinations of alleles may present an advantage in immune response such that they are preserved, or a disadvantage such that recombinants containing an unfavorable combination of alleles does not compete well enough to contribute to future generations.

**Class I molecules**

Made up of an alpha chain \((\alpha)\) of 45 kDa that is a transmembrane protein, and \(\beta_2\text{-microglobulin} (\beta_2\text{m})\) of 12 kDa that is not transmembrane and that associates non-covalently with the \(\alpha\) chain.

The alpha chain consists of three extracellular domains \((\alpha_1, \alpha_2\text{ and }\alpha_3)\), a transmembrane domain and a short cytoplasmic domain encoded as shown in Figure 7-5. The \(\alpha_1\) domain is closest to the membrane and makes the primary contacts with \(\beta_2\text{m}\). Both the \(\alpha_1\) domain and \(\beta_2\text{m}\) are about 100 amino acids in length and have structures similar to the immunoglobulin fold seen in individual immunoglobulin C region domains. The \(\alpha_1\) and \(\alpha_2\) domains fold up together to form a cleft with a \(\beta\) pleated sheet for the floor and alpha helices for the sides.

The mouse H-2K, H-2D and H-2L proteins (see Figure 7-1) are class I proteins. A heterozygous mouse can inherit a different polymorphic allele encoding each protein from both its parents. Thus it can express as many as six different class I proteins. All except a very few cell types in that mouse expresses all six class I molecules on their cell surface.

Humans have three class I regions, each encoding a class I molecule (HLA-A, HLA-B and HLA-C. Thus people heterozygous at all three loci express six different class I molecules on each cell.

(Mice or humans homozygous at the MHC can only express three different class I molecules on each cell.)

**The function of class I molecules is to bind peptides from proteins degraded in the cytoplasm and present them to T cells.**
As shown in Figures 7-10a, 7-12 and 7-13, the peptide is largely extended in the groove and its ends are buried. Cleft is blocked at both ends. Peptide is generally no longer than 8-10 amino acids. It may bulge out in the middle if somewhat long. Binding of peptide is very stable as determined by equilibrium dialysis.

Figure 7-11 shows you examples of the peptides bound by two mouse class I MHC molecules – H-2D^d and H-2K^d. Notice that the so-called “anchor residues” differ for the peptides bound by the two class I molecules. As long as the anchor requirements are satisfied, the rest of the peptide sequence does not matter. The anchor residues fit into the pockets at the ends of the groove of the class I molecule.

Polymorphisms that distinguish different alleles at the class I loci are primarily in the floor of the peptide-binding cleft or in the α-helices at the sides of the cleft. This makes the “shape” of the cleft different, and so the different alleles bind a different set of peptides.

Thus the set of peptides that can be presented by an individual’s class I molecule depends on the alleles he/she inherited from his/her parents. This means that different T cell epitopes may be recognized by each individual. Our class I-specific T cell responses are determined first by the peptides our class I molecules present and second, by the T cell receptors that are available to recognize them.

There are about 1 x 10^5 class I molecules expressed on an average cell. Studies suggest that as many as 2,000 different peptides are present in the grooves of one kind of class I molecule (e.g., H-2D^d), and that from 100 to 4000 copies of any single peptide may be presented per cell. This suggests that as few as 100 copies of a peptide/MHC complex may be a sufficient target for lysis by a cytotoxic T cell that recognizes it.

Each cell of an individual expresses all alleles inherited from his or her parent. Amount expresses varies with cell type. Lymphocytes express the most (5 x 10^5 class I molecules per cell), while liver hepatocytes, muscle and fibroblasts express very little. Neurons, sperm and placenta may not express any.

Class II MHC molecules present peptides derived from proteins taken up from outside the cell – not produced by the cell itself

Class II molecules consist of two chains, each of which is a transmembrane protein. These chains associate non-covalently. 33 KDa α chain, 28 KDa β chain. The membrane-proximal domains of each chain are Ig-like. The membrane-distal domains (α and β) fold up together to form a cleft very much like that seen in class I.

Unlike class I, the cleft of class II molecules is open at each end and can accommodate peptides of 13-18 amino acids. The core binding sequence is 7-10 amino acids in length, and generally consists of an amino acid with an aromatic or hydrophobic side chain at the beginning of the core, and 3 more hydrophobic amino acids at the middle and end of the core.

Remember – class II molecules are expressed only on antigen-presenting cell (APC). A single APC expresses all of the class II chains encoded by its genome, a maximum of:

- 2 alleles of the IA_α locus;
- 2 alleles of the IA_β locus;
- 2 alleles of the IE_α locus; and
- 2 alleles of the IE_β locus.

Since the IA_α chain inherited from one parent can assemble with the IA_β chain inherited from the other parent, and similarly for the IE locus, an APC of a mouse heterozygous at all class II loci will express 8 different class II (Figure 7-16). Thus an APC – particularly one heterozygous at all class II loci – can present a large number of different peptides.

Expression of class II molecules on a cell can be induced from none or a low level to high levels following activation of a cell.
Detailed genomic map of MHC loci

In the mouse, there are about 100 genes in an area of approximately 2 million nucleotides. These consist generally of:

Class I

Classical class I genes (K,D and L)

Non-classical class I genes located 3’ (or downstream) of L. Structurally, these look like classical class I genes, but their tissue distribution is much more limited. These may all have special functions. For example, H-2M is thought to be specialized to present peptides from prokaryotic organisms that grow inside cells such as Mycobacterium tuberculosis and Salmonella typhi and in which the amino-terminal methionine of proteins is formylated. This was suspected because H-2M can present a formylated peptide from a protein (NADH dehydrogenase) encoded by the mitochondrial genome.

Class II region

Classical class II genes (IA and IE)

Non-classical class II genes with limited polymorphism and a different tissue distribution are encoded 5’ (or upstream) of the classical class II genes. We generally do not know what these do. In humans, the product of the HLA-DM class II gene is involved in loading peptides from degraded external antigens onto class II molecules.

Special genes involved in antigen processing (Chapter 8):

- TAP1 and TAP2 genes that act as transporters to carry peptides from proteins degraded in the cytoplasm into the endoplasmic reticulum for binding to class I molecules;
- LMP2 and LMP7 proteins that form part of the proteasome, the large complex of proteases that degrades proteins produced in the cytoplasm into peptides that can be presented by class I MHC molecules.

Regulation of MHC expression

Nucleic acid sequences located in either 3’ or 5’ flanking regions are targets for positive or negative regulation of MHC genes.

Mutations in these sequences or in transcription factor or other regulatory proteins that recognize them can lead to immunodeficiency.

For example, in the human bare lymphocyte syndrome, defects in the class II transcriptional activators RFX and CIITA result in failure to express class II molecules, resulting in severe immunodeficiency.

Cytokines are known to have effects on expression of MHC molecules and other immune-related molecules. For example:

- Interferons α, β and γ as well as tumor necrosis factor I (TNF-I) all signal cells to increase expression of class I MHC α chain, β2-microglobulin, LMP-2 and LMP-7, and TAP1 and TAP2.
- Interferon γ also increases class II MHC expression on a variety of cells, some of which are professional APCs (e.g., macrophages, dendritic cells) and others of which are not (e.g., vascular endothelial cells, pancreatic beta cells).

Sometimes cytokines have opposite effects on class II expression: for example, IL-4 increases class II on B cells, and IFN-α leads to a decrease. We’ll see more examples of complex regulation by cytokines in later chapters.
Finally, some viruses have adapted to decrease anti-viral immune responses by altering expression of MHC molecules by their hosts:

For example, human cytomegalovirus (CMV), a sexually-transmitted virus, encodes and produces a protein that binds human β2-microglobulin, thereby preventing class I molecule assembly. Cells with less class I to present viral peptides are more difficult for cytotoxic T lymphocytes to kill.

**The MHC can often dictate the strength or weakness of an individual’s immune response to particular antigens or infectious organisms**

The alleles at the classical class II play a role in determining the strength of immune response to external stimuli. Linkage of strong or weak antibody responses to the MHC was first shown by Baruj Benacerraf who studied antibody responses of inbred strains of guinea pigs to complex synthetic polypeptide antigens. Later it was shown by Hugh McDevitt, working with inbred and congenic mouse strains, that the strength of the response mapped to regions of the MHC called IA or IE, later shown to encode class II genes.

Failure to respond to certain antigens in a MHC-linked fashion has two possible mechanistic explanations, both of which have shown to be causal in different cases:

One is referred to as **determinant selection**, and means that a poorly responding strain is unable to present any peptide derived from the antigen to the T cells necessary for the response to be initiated. This could be due to polymorphism in class II molecules or in proteins that generate (proteasome) or transport (TAP) the peptides.

The other is called **hole-in the repertoire**. In this case, no T cells with T cell receptors that can bind peptide/MHC complexes derived from the immunizing antigen exist in the individual. This occurs because peptides identical or similar to the immunizing antigen can be generated from the individual’s own self-proteins. These bind to his/her own MHC molecules and delete reactive T cells in the thymus as being autoreactive. No such T cells ever get to be members of the mature T cell repertoire.

Both situations could lead to increased susceptibility to a pathogen. Inability to mount a response to key epitopes of the pathogen could result in a severe disadvantage.

**MHC and disease susceptibility**

A number of inherited human diseases are associated with particular MHC alleles. These include autoimmune diseases, some viral diseases, complement deficiencies, neurologic disorders and allergies. This has led to the concept of **relative risk** which is expressed as follows:

\[
RR = \frac{[\text{allele}^+ \text{ individuals}] [\text{allele}^- \text{ individuals}] \text{ in diseased population}}{[\text{allele}^+ \text{ individuals}] [\text{allele}^- \text{ individuals}] \text{ in control population}}
\]

As shown in Table 7-3, a RR = 1 means there is no association with the allele.

If RR is > 1, there is some association. For example, the presence of the HLA-B27 allele renders an individual 90 times more likely to have ankylosing spondylitis, an inflammatory disease of vertebral joints involving the destruction of cartilage.

This does not necessarily mean that the allele caused the disease. For example, consider HLA-A3 and HLA-B14. Individuals homozygous for both have an RR of 90 for hereditary hemochromatosis, a disease in which patients retain 5-10 times more iron than normal. This allele is not due to the alleles themselves. A mutation in **HLA-H**, a gene close to HLA-A that encodes a non-classical class I MHC molecule, prevents it from associating with β2-microglobulin. Result is that there is no HLA-H on the surface of cells in the stomach, intestines and liver. Evidence suggests that HLA-H plays a role in iron uptake.
What might it mean if RR is < 1?

Why would a disadvantageous allele persist in the population and not die out? The trait is recessive and heterozygotes are therefore normal. They serve as carriers. Also, the disease manifests itself at age 40-50 and therefore has a minimal effect on reproductive performance.