The Major Histocompatibility Complex

Chapter 6

WHAT'S IN THIS CHAPTER?

Near or far, hiddenly,
To each other linked are,
That thou canst not stir a flower
Without troubling a star.

Francis Thompson

As was introduced in Chapter 2, recognition of antigen by αβ T cells is more complex than antigen recognition by B cells. While the BCR binds directly to a unitary epitope on a pathogen or foreign macromolecule, another host cell must "present" antigens derived from the pathogen to T cells. The epitope recognized by an αβ T cell's TCR is a peptide derived from a protein antigen displayed on a cell surface molecule encoded by one of the genes of the **major histocompatibility complex (MHC)**.

A. Overview of the Major Histocompatibility Complex

Proteins encoded by the MHC were originally discovered in the 1930s during studies of tissue rejection in transplantation experiments. These proteins were therefore named for their association with **histocompatibility** (*histo*, meaning "tissue," and *compatibility*, meaning "getting along"). The genes controlling the histocompatibility of tissue transplantation were localized to a large genetic region containing multiple loci; hence, the term "complex." Moreover, the proteins encoded by these genes were found to have dramatic effects on histocompatibility. To distinguish these proteins from other molecules (encoded elsewhere in the genome) that had relatively minor effects on histocompatibility, these molecules were called the "major" histocompatibility molecules. Thus, the genes encoding these proteins were dubbed the "major histocompatibility complex" (MHC) genes. Soon after, it was discovered that MHC-controlled rejection of transplanted tissue was due to the mounting by the transplant recipient of an immune response against the donated cells (see Ch. 17). Although this finding implied that MHC gene products were directly involved in immune responses, it took several more decades for immunologists to define the normal physiological role of MHC-encoded proteins in presenting antigenic peptides to T cells.

MAK: Primer to the Immune Response. http://dx.doi.org/10.1016/B978-0-12-385245-8.00006-6

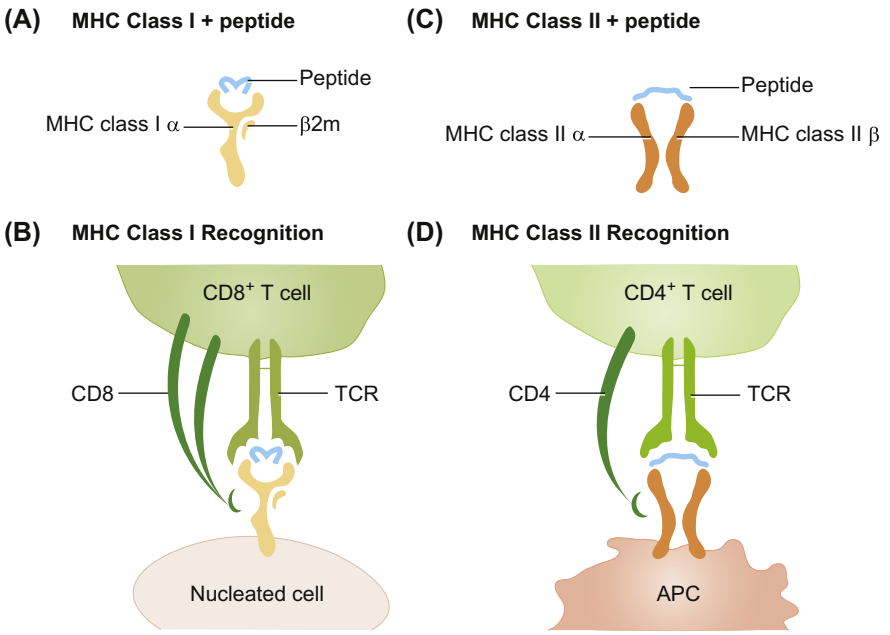


Fig. 6-1
Recognition of MHC Class I and II Molecules by T Cells

(A) An MHC class I protein contains a large transmembrane α chain associated with the smaller β2m chain. **(B)** The TCR of a $CD8^+$ T cell binds to a peptide-MHC class I complex on a nucleated host cell, while its CD8 coreceptor binds to another site on the same MHC class I molecule. **(C)** An MHC class II protein contains two large transmembrane chains: α and β. **(D)** The TCR of a $CD4^+$ T cell binds to a peptide-MHC class II complex on an APC, while its CD4 coreceptor binds to another site on the same MHC class II molecule.

TABLE 6-1 Chromosomal Location of MHC Class I and II Genes

Protein Encoded	Chromosome	
	Human	Mouse
MHC class I α chain	6	17
β2-microglobulin	15	2
MHC class II α chain	6	17
MHC class II β chain	6	17

The MHC-encoded proteins that are involved in most instances of antigen recognition by T cells are the **MHC class I** and **MHC class II** molecules. The TCRs of $CD8^+$ T cells recognize peptides bound to MHC class I, while the TCRs of $CD4^+$ T cells recognize peptides bound to MHC class II (Fig. 6-1). As is described in detail in Chapter 8, the CD8 coreceptor of $CD8^+$ T cells also binds to MHC class I, while the CD4 coreceptor of $CD4^+$ T cells binds to MHC class II. The MHC class I protein is a heterodimer consisting of a large transmembrane α chain non-covalently linked to a small non-transmembrane chain called **β2-microglobulin (β2m)**. The MHC class I α chain is encoded within the MHC, but β2m is not (Table 6-1). The MHC class II protein is composed of an α chain and a slightly smaller β chain, both of which are transmembrane proteins encoded by genes in the MHC. Despite this difference in composition, the tertiary structures of MHC class I and class II molecules are highly similar, apart from the peptide-binding groove. While almost all nucleated cells express MHC class I, only the few cell types that function as APCs (including DCs, macrophages and B cells) express MHC class II. Thus, almost any cell can serve as a target cell and present antigen to CTLs derived from $CD8^+$ Tc cells, but only APCs can activate $CD4^+$ Th cells.

I. HLA Complex

In the human genome, the MHC is called the **HLA complex** (for human leukocyte antigen complex). The HLA complex covers about 3500 kb on chromosome 6 and contains 12 major regions, as shown in Figure 6-2A. Each region contains dozens of genes, only some of which are functional and many of which do not appear to be involved in antigen presentation. The HLA-A, HLA-B and HLA-C regions are all MHC class I regions. Each contains a single functional gene encoding a human MHC class I α chain.

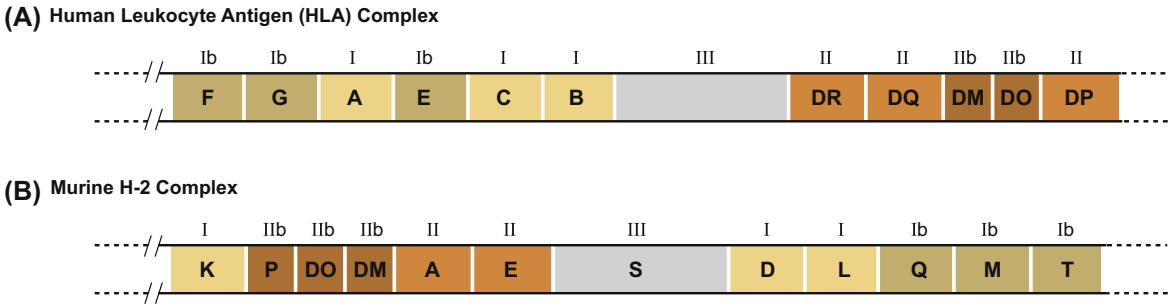


Fig. 6-2
General Organization of the MHC in Humans and Mice

Schematic representation of chromosomal regions in which MHC class I, Ib, II, IIb and III genes are found. **(A)** The human leukocyte antigen (HLA) complex on chromosome 6. Note that regions in the HLA complex containing MHC class III genes do not have letter names. **(B)** The murine MHC (H-2) complex on chromosome 17. *[Source:* http://imgt.cines.fr/*.]*

The DP, DQ and DR regions are all MHC class II regions. Each contains multiple functional genes encoding both MHC class II α and β chains. The single genes within each of the HLA-E, -F and -G regions encode **MHC class Ib** proteins, while several genes in the DM and DO regions encode **MHC class IIb** proteins. MHC class Ib and IIb proteins structurally resemble MHC class I and II proteins, respectively, but are not directly involved in routine antigen presentation to T cells. MHC class Ib and IIb proteins are therefore considered to be "non-classical" MHC molecules (see **Box 6-1**). The **MHC class III** region is not known to encode any peptide-binding presentation molecules but contains many genes relevant to immune responses, including those encoding complement components, HSPs and the cytokines TNF and **lymphotoxin (LT)**.

Box 6-1 Human Non-classical and MHC-like Genes and Proteins

The "non-classical" MHC class Ib and IIb molecules resemble the classical MHC proteins in structure but generally do not present peptides to αβ T cells (see table below). The class Ib genes in humans are located in the HLA-E, -F and -G regions. Some class Ib gene products are secreted (unlike the products of classical MHC class I genes), while others are membrane-bound. Two MHC class Ib proteins called HLA-E and HLA-F may function in antigen presentation to γδ T cells (see Ch. 11). HLA-G is expressed in placental cells during fetal development and may contribute to the prevention of maternal immune responses against the fetus (see Ch. 10). A gene called HFE, which is located in the HLA-E region, encodes an MHC class I-like protein that associates with β2m but does not have a peptide-binding groove. HFE appears to be involved in iron absorption, such that when HFE is defective, excessive iron is deposited in various organs. Two MHC class Ib proteins called **MICA** and **MICB** closely resemble MHC class I molecules in structure but are stress-induced molecules. The binding of MICA to a particular receptor on NK and γδ T cells can stimulate these cells (see Ch. 11). The human MHC class IIb proteins are located in the DO and DM regions. The HLA-DM protein regulates the loading of antigenic peptides onto MHC class II molecules (see Ch. 7), while the HLA-DO protein inhibits HLA-DM activity.

The **CD1** proteins are encoded *outside* the MHC loci but share certain structural similarities and functions with classical MHC molecules. These "MCH-like" CD1 proteins associate with β2m but have binding grooves that are very hydrophobic. This type of groove preferentially binds fragments of lipid and glycolipid antigens. Certain T lineage cells can be activated by this non-peptidic form of antigen presentation (see Ch. 7).

	Class I	Class II	Class Ib	Class IIb	Class III	MHC-like
Gene encoded in MHC region	Yes	Yes	Yes	Yes	Yes	No
Polypeptides	Class I α plus β2m	Class II α plus Class II β	Class I α-like plus β2m	Class II α-like plus Class II β-like	Neither class I nor class II chains	Non-MHC chains plus β2m

	Class I	Class II	Class Ib	Class IIb	Class III	MHC-like
Tissue expression	Almost ubiquitous	APCs	Restricted	APCs	Almost ubiquitous	Restricted
Soluble form?	Very rare	No	Some	Some	Yes	Some
Polymorphism	Extreme	Extreme	Limited	None	None	None
Function	Peptide presentation to $CD8^+$ T cells	Peptide presentation to $CD4^+$ T cells	Stimulation of γδ T or NK cells; fetus protection; iron absorption	Peptide loading of MHC class II	Complement components, inflammatory cytokines, heat shock and stress proteins	Lipid antigen presentation to T lineage cells
Examples	HLA-A HLA-B HLA-C	HLA-DP HLA-DQ HLA-DR	HLA-E HLA-F HLA-G HFE MICA MICB	HLA-DM HLA-DO	C4 TNF HSP70	CD1 isoforms

II. H-2 Complex

In the mouse genome, the MHC is known as the **H-2 complex**. The H-2 complex is spread over 3000 kb on chromosome 17 and contains 12 major regions, as shown in **Figure 6-2B**. The K, D and L regions contain single functional genes that encode mouse MHC class I α chains. The A and E regions each contain a single functional gene encoding an MHC class II α chain, and one or more functional genes encoding an MHC class II β chain. The S region of the H-2 complex contains genes encoding the MHC class III proteins, again including complement proteins, HSPs, TNF and LT. The Q, T and M regions of the H-2 complex contain genes encoding class Ib proteins, whereas class IIb proteins are encoded by genes in the P, DO and DM regions.

B. MHC Class I and Class II Proteins

The MHC class I and class II proteins are heterodimeric molecules composed of an extracellular N-terminal peptide-binding region, an Ig domain-containing extracellular region, a hydrophobic transmembrane region, and a short C-terminal cytoplasmic region. The structure of the peptide-binding region in both MHC class I and class II molecules is such that the affinity of an MHC molecule for peptide is much lower than that of an antibody for its cognate antigen. This relaxed binding is a necessity if a given MHC molecule is to carry out its task of presenting a wide range of peptides for T cell perusal. It should also be noted that a given peptide may be capable of binding to different MHC class I or class II molecules, a phenomenon known as "promiscuous binding."

I. MHC Class I Proteins

Most nucleated cells sport a mixed population of MHC class I proteins. The peptides bound by these MHC class I molecules are generally of **endogenous** origin; that is, they are derived from the degradation of proteins synthesized within the cell. The vast majority of these peptides will be "self" in nature, because most proteins routinely produced within a cell at any one time are of host origin (as opposed to proteins of

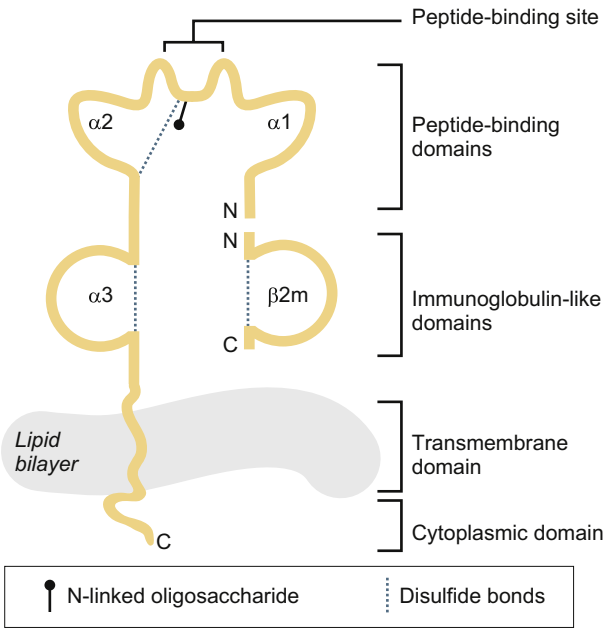


Fig. 6-3
Structure of the MHC Class I Protein

Schematic representation showing component chains and domains of an MHC class I protein and its position in a host cell membrane. N, amino-terminus; C, carboxy-terminus.

non-self origin, such as those generated during a viral infection). The MHC class I molecule does not discriminate among "self" and "non-self" peptides; that job is left to the TCRs of $CD8^+$ T cells. Self peptide–MHC complexes do not trigger an immune response because T cells with the corresponding specificity are generally absent from the T cell repertoire due to the establishment of central tolerance (see Ch. 9). In contrast, non-self peptides complexed to MHC class I are recognized and trigger $CD8^+$ T cell activation.

How peptides are produced from antigens and loaded into the peptide-binding grooves of MHC class I or II molecules is described in Chapter 7.

i) MHC Class I Component Polypeptides

In both mice and humans, MHC class I α chains are glycoproteins of about 44 kDa and contain three extracellular globular domains (**Fig. 6-3**). Domains α1 and α2 at the N-terminal end of the chain non-covalently pair with each other to form the peptide-binding site, while the Ig-like α3 domain associates non-covalently with the β2m polypeptide. The α chain also supplies the transmembrane domain and the cytoplasmic domain. The α1 domain maintains its shape without disulfide linkage, but the α2 and α3 domains each have an internal disulfide bond. The other partner in the MHC class I molecule, the β2m protein, is a non-transmembrane polypeptide of about 12 kDa. β2m resembles a single Ig-like domain and, through its association with the MHC class I α3 domain, helps to maintain the overall conformation of the MHC class I molecule. Indeed, the binding of β2m to the MHC class I α chain soon after protein synthesis in the ER is essential for the transportation of the complete heterodimer to the cell surface.

ii) MHC Class I Peptide-Binding Site

The groove-like peptide-binding site of the MHC class I molecule is relatively small. As a result, MHC molecules cannot recognize large native antigens. Rather, antigens must be processed into small peptides that can fit into the MHC groove before they can be presented to T cells. It has been estimated that each MHC class I molecule has the ability to bind to several hundred different peptides with moderately high affinity but captures only one peptide at a time.

The MHC class I peptide-binding groove is formed by the juxtaposition and interaction of the α1 and α2 domains of the α chain. The β2m chain contributes by interacting with the amino acids in α1 and α2 that form the floor of the groove. These interactions are strengthened, and the entire MHC class I structure is stabilized when the groove is occupied by a peptide of 8–10 amino acids. The peptide is

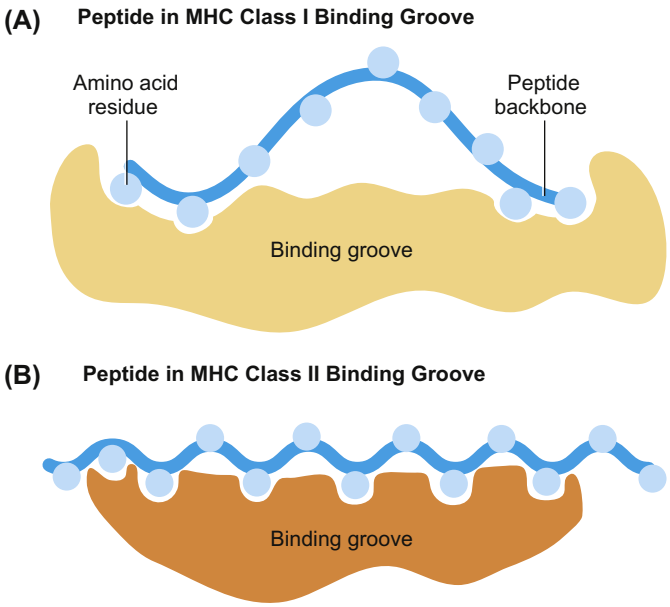


Fig. 6-4
MHC Peptide-Binding Sites

Amino acid positioning within the peptide-binding grooves of MHC class I and II proteins is shown. Blue circles represent individual amino acids of the antigenic peptides. **(A)** Peptides binding in an MHC class I groove are usually 8–10 amino acids long. The residues at each end of the peptide are anchored in the groove. **(B)** Peptides binding in an MHC class II groove are usually 13–18 amino acids long. The residues at each end of the peptide may overhang the groove and do not serve as anchor residues.

held in place in the groove by interactions between specific amino acids of the α1 and α2 domains and conserved "anchor residues" located in the N- and C-termini of the peptide. The peptide anchor residues point "down" into the groove, while the central peptide residues project "up" toward the TCR (**Fig. 6-4A**). A sufficient degree of conformational flexibility exists such that peptides of widely varying amino acid sequence in the region between the anchor residues can occupy the groove. The ends of the MHC class I groove are closed, which means peptides larger than 8–10 amino acids can fit in only if their central residues can bulge upward out of the groove.

II. MHC Class II Proteins

As mentioned earlier, MHC class II molecules are found almost exclusively on APCs. The peptides bound by MHC class II are of **exogenous** origin; that is, derived from the degradation of proteins that have entered the cell from the exterior via either phagocytosis or receptor-mediated endocytosis. Because APCs also capture and digest spent host proteins, the vast majority of peptides presented on MHC class II molecules are "self" and do not trigger $CD4^+$ T cell activation because these specificities have been removed from the Th cell repertoire by the establishment of central tolerance. A Th response is induced when an APC presents a non-self peptide bound to MHC class II.

i) MHC Class II Component Polypeptides

In both humans and mice, the α and β chains of MHC class II proteins are glycoproteins of similar size and structure (24–32 and 29–31 kDa, respectively). Both chains contain an N-terminal extracellular domain, an extracellular Ig-like domain, a hydrophobic transmembrane domain, and a short cytoplasmic tail (**Fig. 6-5**). The peptide-binding region is made up of the N-terminal α1 and β1 domains of the α and β chains, respectively. The α2 and β2 domains form globular loops that are homologous to the Ig fold but are not involved in peptide binding.

ii) MHC Class II Peptide-Binding Site

The MHC class II peptide-binding groove is similar in overall structure to that of MHC class I molecules (**Fig. 6-4B**). However, the ends of the MHC class II groove are open, permitting the binding of much longer peptides (up to 30 amino acids). Nevertheless, the majority of peptides found in MHC class II grooves are 13–18 amino acids long. The open ends of the MHC class II groove also mean that binding does not depend on

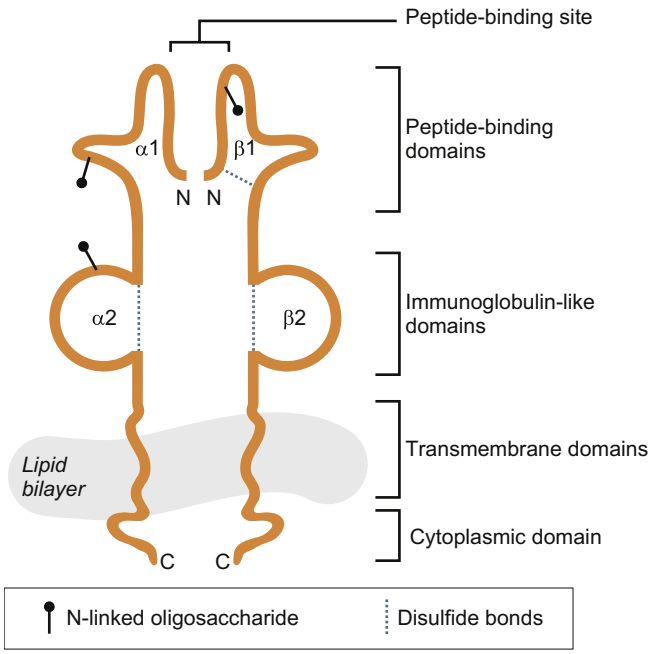


Fig. 6-5
Structure of the MHC Class II Protein

Schematic representation showing component chains and domains of an MHC class II protein and its position in the APC membrane. N, amino-terminus; C, carboxy-terminus.

conserved anchor residues at the ends of the peptides but is instead mediated by hydrogen bonding between the peptide backbone and the sidechains of certain MHC amino acids. Researchers have found that antigenic peptides that are successfully bound to the floor of the MHC class II groove possess a particular conserved secondary structure (resembling a polyproline chain) in the portion of the peptide that aligns with critical acidic MHC residues located in the middle of the groove. As a result of this conformational requirement, MHC class II proteins generally bind a narrower range of proteins than do MHC class I proteins.

III. X-Ray Crystallography of MHC Class I and II Molecules

Much of the information on how MHC class I and II molecules bind to peptides has come from X-ray studies of crystallized pMHC complexes. **Plate 6-1** shows the crystal structures of the carbon backbones of the extracellular regions and peptide-binding grooves of murine MHC class I and MHC class II molecules. The similarity of their tertiary structures can be clearly seen. Analyses of such MHC crystal structures have shown that water plays an important role in peptide–MHC binding. The fit of the peptide in the groove is tightened when water molecules fill any gaps in the complex.

NOTE: As X-ray crystallography techniques become more and more refined, the number of macromolecular 3-D crystal structures available in public databases grows. A searchable database of protein structures is maintained by the U.S. National Center for Biotechnology Information (NCBI) and can be found at the website http://www.ncbi.nlm.nih.gov/sites/structure. At this site, the structures of hundreds of pMHC complexes can be viewed in 3-D.

C. MHC Class I and Class II Genes

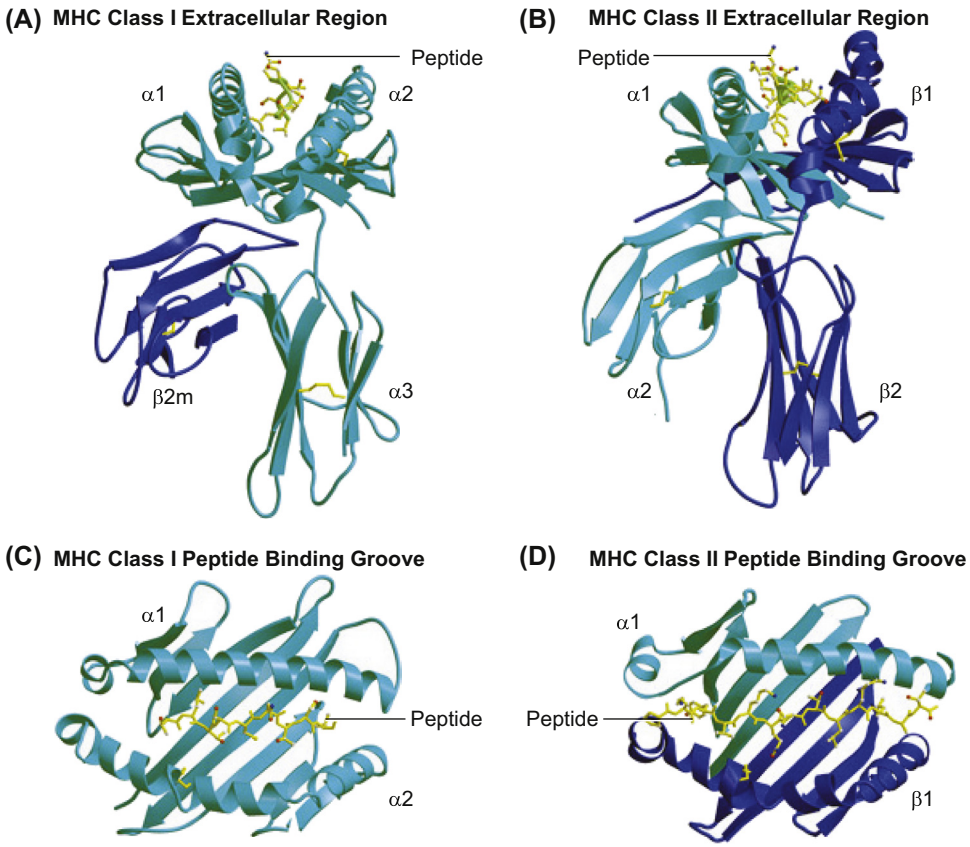
I. Polygenicity of MHC Class I and II Genes

Most proteins in our bodies are unique; that is, there is only one functional gene in the genome that encodes a protein carrying out that particular function. The MHC genes are unusual in that, due to gene duplication during evolution, two to three separate, functional genes encoding the same type of MHC class I or II polypeptide exist. This

Plate 6-1

X-Ray Crystal Structures of MHC Class I and II Molecules in the Mouse

Crystal structures showing the carbon backbone of murine MHC class I and II molecules. **(A)** and **(B)** show peptides bound to the extracellular regions of MHC class I or II, respectively. **(C)** and **(D)** show the view looking down at the peptide in the peptide-binding groove of MHC class I or II, respectively. *[Reproduced by permission of Bjorkman, P.J. (1977). MHC restriction in three dimensions: A view of T cell receptor/ligand interactions.* Cell *9, 167–170.]*



phenomenon is called **polygenicity**. These genes are named for their region of location and the chain they specify. For example, the HLA includes three loci, HLA-A, -B and -C, that all encode the same type of polypeptide: an MHC class I α chain. Similarly, genes giving rise to MHC class II α chains can be found in the HLA-DP, -DQ and -DR regions; these genes are called DPA, DQA and DRA genes, respectively. Also in each of the HLA-DP, -DQ and -DR regions are separate genes that encode MHC class II β chains; these are called DPB, DQB and DRB genes, respectively.

The MHC class II loci show further polygenicity in that each of the HLA-DP, -DQ and -DR regions may have more than one α chain gene and more than one β chain gene. For example, within the HLA-DP region, there are two genes that could encode MHC class II α chains, DPA1 and DPA2, and two genes that could encode MHC class II β chains, DPB1 and DPB2. However, only the DPA1 and DPB1 genes are functional. Similarly, the HLA-DQ region contains the DQA1 and DQA2 genes that could encode MHC class II α chains, and the DQB1, DQB2 and DQB3 genes that could encode MHC class II β chains. However, only DQA1 and DQB1 are functional. In the HLA-DR region, a single gene designated DRA encodes the DR α chain and is functional in all humans. In contrast, not every individual carries the same number of DRB loci on his/her chromosomes. Nine different DRB genes have been identified, designated DRB1 to DRB9. While every individual has the DRB1 and DRB9 loci, different individuals may also have one or more DRB loci selected from among the DRB2 to DRB8 genes. However, only DRB1, DRB3, DRB4 and DRB5 are functional and encode DR β chains.

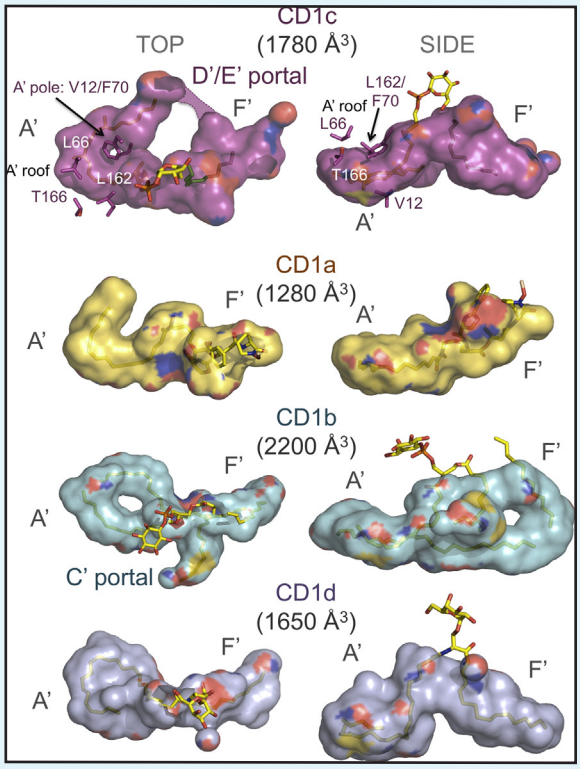
For unknown reasons, an α chain derived from a DP region gene almost always combines with a β chain derived from the DP region (and not from the DQ or DR regions) to form a complete MHC class II molecule. Similarly, a DQ α chain combines with a DQ β chain and a DR α chain with a DR β chain. Only very rarely do mixed MHC class II molecules such as HLA-DRA/DQB occur. **Figure 6-6A** illustrates how the products of the HLA loci can come together to form complete human MHC molecules.

Focus on Relevant Research

"The 2.5 Å Structure of CD1c in Complex with a Mycobacterial Lipid Reveals an Open Groove Ideally Suited for Diverse Antigen Presentation." by Scharf, L., Li, N.S., Hawk, A.J., Garzón, D., Zhang, T., Fox, L.M., Kazen, A.R., Shah, S., Haddadian, E.J., Gumperz, J.E., Saghatelian, A., Faraldo-Gómez, J.D., Meredith, S.C., Piccirilli, J.A., and Adams, E.J. (2010) *Immunity 33,* 853–862.

Nearly one-third of the world's population is infected with the intracellular bacterium *Mycobacterium tuberculosis*, the causative agent of tuberculosis (TB), and nearly 2 million people die worldwide each year from this infection. It is therefore no surprise that much research has been aimed at understanding the immune responses associated with TB. One hallmark feature of this bacterium is its cell wall, which contains unusual and abundant lipids. Scientists have known for some time that T cell-mediated responses are important for protection against TB but that these responses are not triggered by conventional pMHC complexes presented by MHC class I and class II molecules. Instead, anti-TB T cell responses are mounted against bacterial lipids presented by CD1 molecules, which are "MHC-like" proteins encoded by a gene outside the MHC (see Box 6-1). Structurally, CD1 molecules have evolved to possess deep, narrow binding grooves that excel at accommodating and anchoring the hydrophobic alkyl chains of lipid molecules.

In this article, Scharf *et al.* report on the precise way in which CD1c, a particularly interesting CD1 isoform, displays lipid antigens to T cells. The authors use the powerful technique of X-ray diffraction crystallography to construct a three-dimensional (3-D) model showing how lipids derived from *M. tuberculosis* are displayed in the open groove of the CD1c molecule. They then use this information to explain why CD1c is uniquely suited to displaying a wide variety of bacterial lipids. Moreover, they determine the lipid-CD1c interaction to a resolution of 2.5 angstroms (2.5 Å), arguably the finest level of detail of macromolecular structure that can be determined experimentally. In Figure 3 from the article (shown here), we see 3-D representations of the space occupied by the antigen-binding cavities of CD1c (magenta), CD1a (yellow), CD1b (cyan) and CD1d (light purple), with the red, blue and yellow areas representing the contributions of oxygen, nitrogen and sulfur atoms, respectively, to each cavity surface. The volume of each cavity is given in cubic angstroms (Å^3), and two major pockets can be identified (A' and F'). The small red and yellow stick models represent lipid antigens bound inside the cavities. In the case of CD1c, we also see amino acid residues that form conserved cavity features (L66, L162, T166) and an important inter-residue contact (V12/F70). Using these types of models to better understand how APCs present lipid antigens to T cells via CD1 isoforms may enhance our ability to create more effective vaccines against important pathogens like *M. tuberculosis*.



Polygenicity also occurs in the mouse H-2 complex. Two loci, H-2K and H-2D, contain single genes encoding MHC class I α chains. MHC class II α chains are encoded by one functional gene called Aa (or Aα) within the A region of the mouse H-2, as well as by the Ea (or Eα) gene in the E region. Similarly, MHC class II β chains are encoded by the Ab (or Aβ) gene in the H-2A region and the Eb (or Eβ) gene in the H-2E region. **Figure 6-6B** shows examples of how products of the H-2 complex give rise to complete murine MHC class I and II molecules. Again, an α chain derived from an A region gene almost always combines with a β chain derived from the A region, and not with an E region β chain (and vice versa).

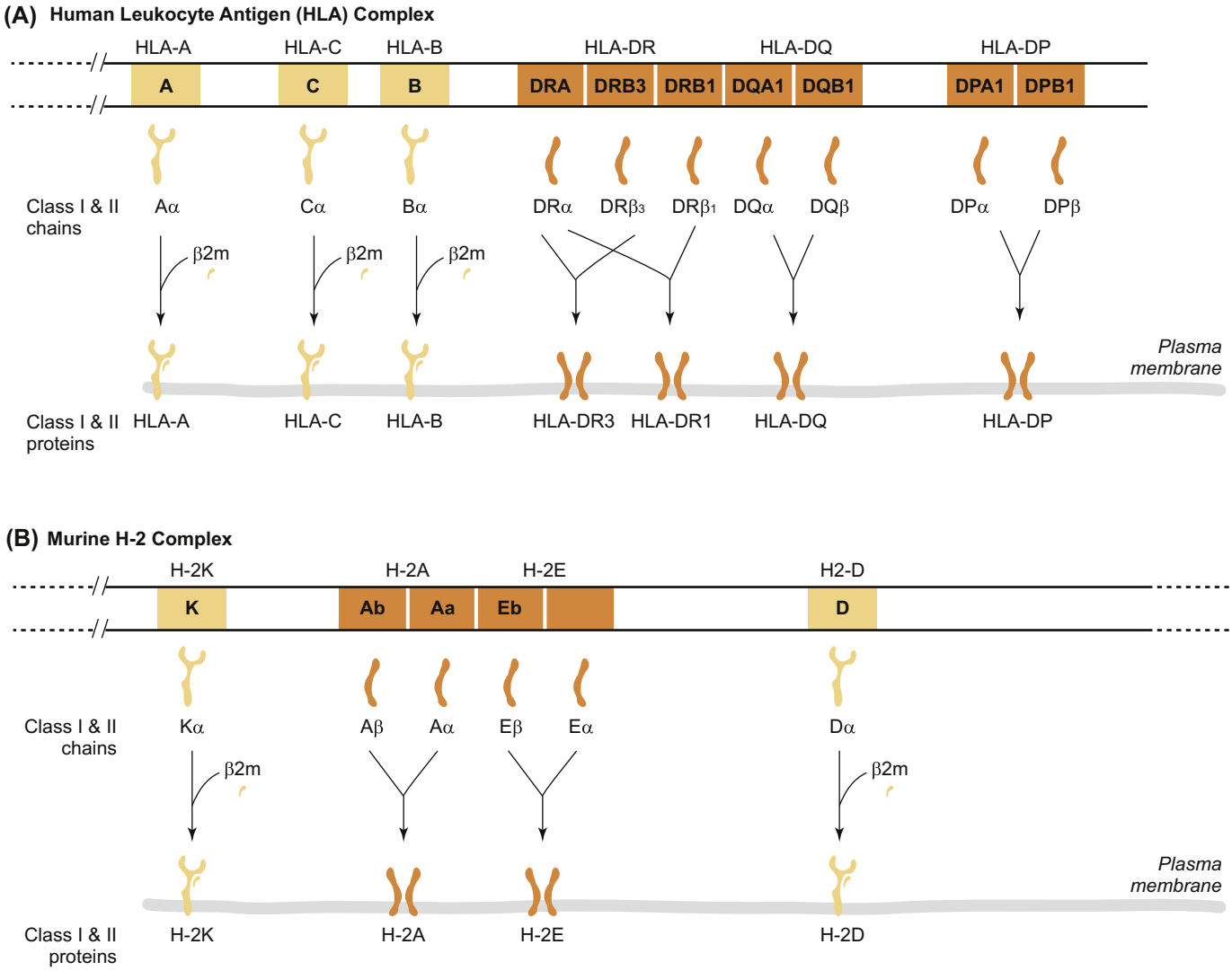


Fig. 6-6
Examples of Polygenicity in the MHC Loci

Multiple genes encode MHC class I and II proteins in the **(A)** HLA and **(B)** H-2 complexes. Examples of how chains derived from each locus can combine to form complete MHC heterodimers are shown. For simplicity, the MHC class Ib, IIb and III loci are not shown.

II. Polymorphism of MHC Class I and II Genes

The vast majority (>90%) of vertebrate genes are **monomorphic**; that is, almost all individuals in the species share the same nucleotide sequence at that locus. In contrast, the MHC loci exhibit extreme polymorphism. **Polymorphism** is the existence in a species of several different alleles at one genetic locus. **Alleles** are slightly different nucleotide sequences of a gene; the protein products of alleles have the same function. For example, close to 1700 alleles have been identified for the HLA-A gene, over 2000 for HLA-B, and more than 1200 for HLA-C (Table 6-2). A functional MHC class I molecule can consist of the protein product of any one of these HLA-A, -B or -C alleles associated with the invariant β2m chain. Multiple alleles also exist for the MHC class II genes, so that the product of any DPA allele can combine with the product of any DPB allele (and DQA with DQB, and DRA with DRB) to form a functional MHC class II protein. The degree of sequence variation among MHC alleles can be astonishing: differences of as many as 56 amino acids have been identified between individual alleles. Not surprisingly, this amino acid variation is concentrated in the peptide-binding site of

TABLE 6-2 Numbers of HLA Alleles

	Number of Alleles*
MHC Class I Genes	
HLA-A	1698
HLA-B	2271
HLA-C	1213
MHC Class II Genes	
HLA-DPA1	32
HLA-DPB1	149
HLA-DQA1	44
HLA-DQB1	158
HLA-DRA	7
HLA-DRB	1074

*Data are from the Immunogenetics HLA (IMGT HLA) database maintained by the European Bioinformatics Institute at EMBL (http://www.ebi.ac.uk/imgt/hla/) and represent alleles reported as of September 2011.

New MHC alleles are discovered every day. Up-to-date numbers of MHC alleles can be found in the IMGT/HLA database maintained by the European Bioinformatics Institute at EMBL.

the MHC protein. In MHC class I α chains, most of the polymorphism is localized in the α1 and α2 domains. The α3 domain is less polymorphic and more Ig-like, and the transmembrane and cytoplasmic domains are more conserved than any of the α domains. The β2m protein exhibits almost no polymorphism within a species or variation among species. In the case of MHC class II molecules, polymorphic variation among alleles is found in the α1 and β1 domains that constitute the peptide-binding site. Again, the transmembrane and cytoplasmic domains are highly conserved.

Due to the high level of polymorphism in the HLA, humans (which are outbred) are generally *heterozygous* at their MHC loci (have different alleles on the maternal and paternal chromosomes). In contrast, experimental mice (which have been repeatedly inbred to create pure strains) are *homozygous* for any given MHC gene (have the same allele on both the maternal and paternal chromosomes). In addition, in outbred populations, two individuals are very likely to have different nucleotide sequences at each HLA locus. These two individuals are said to be **allogeneic** to each other at their MHC loci (*allo*, meaning "other"). In an inbred population, not only is each individual homozygous at each MHC locus, but all individuals in the population express the same MHC allele at a given locus. Such inbred animals express exactly the same spectrum of MHC molecules and are said to be **syngeneic** at their MHC loci (*syn*, meaning "same").

III. Codominance of MHC Expression

The polygenicity and polymorphism of the MHC genes underlie the vast diversity of MHC molecules expressed within an outbred population. In an individual, the breadth of MHC diversity is increased by the fact that, at *each* MHC locus, the gene on both chromosomes is expressed independently, or **codominantly**. In other words, when a given MHC locus is expressed in an individual, the genes on *both* the maternal and paternal chromosomes produce the corresponding proteins. For example, in an individual heterozygous at the HLA-A locus, there are two MHC class I α chains produced (maternal and paternal) that can combine with β2m to form two different HLA-A proteins. Similarly, for an MHC class II molecule such as HLA-DP, two different α chains and two different β chains are produced that can combine to form four different HLA-DP proteins. For a locus such as HLA-DR, which comprises more than one DRB gene in most individuals, the number of possible HLA-DR heterodimers is much higher. **Figure 6-7** illustrates the net effects of MHC polygenicity, polymorphism and codominant expression in an outbred individual. A nucleated host cell features a wide spectrum of MHC class I molecules on its surface, whereas a typical APC expresses

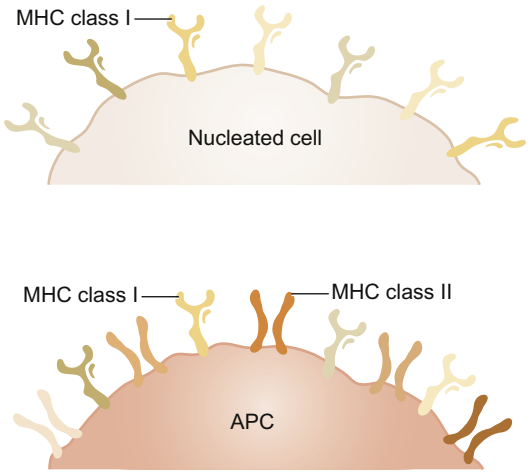


Fig. 6-7
Spectrum of MHC Class I and II Expression in an Outbred Individual

The existence of multiple MHC class I and class II loci, combined with high levels of polymorphism at each locus as well as codominant expression of maternal and paternal alleles at each locus, results in a wide spectrum of MHC molecules being present on the cells of any given individual. All nucleated cells express MHC class I proteins, but only APCs express both MHC class I and II molecules.

multiple types of MHC class II molecules as well as multiple types of MHC class I molecules.

IV. MHC Haplotypes

The MHC loci are closely linked, meaning that the specific set of alleles for all MHC loci on a single chromosome is usually passed on to the next generation as an intact block of DNA. This set of alleles is called a **haplotype**, and any individual inherits two haplotypes from his/her parents: the MHC block on the paternal chromosome and the MHC block on the maternal chromosome. In an outbred population, the high degree of polymorphism of the MHC can lead to great variation in haplotypes among unrelated individuals. Interestingly, within the human population, researchers have identified over 30 *ancestral haplotypes* that are shared not only within a single family but also among a large number of families. These ancestral haplotypes are thought to have originated in "founder" populations that settled in diverse geographic regions, so that a given ancestral haplotype is often associated with a particular ethnic background. For example, one ancestral haplotype is found predominantly among Basques and Sardinians, while another is specific to Eastern European Jews, and a third is exclusive to Southeast Asians.

In inbred mice, both parents have the same allele at each MHC locus, the maternal and paternal haplotypes are the same, and all offspring inherit the same single haplotype on both chromosomes. Immunologists frequently use a single term "short form" to indicate the haplotype of a particular strain. For example, the haplotype of the C57BL/6 mouse strain is denoted "H-2^b", where the "b" in H-2^b means that allele number 12 is present at the K locus, allele number 74 is at the Aβ locus, allele number 3 is at the Aα locus, allele number 18 is at the Eβ locus, and so on. In contrast, the CBA mouse strain has a haplotype of "H-2^k", where allele number 3 is present at the K locus, allele number 22 at Aβ, and so on. More detailed short forms can be used to indicate specific alleles in a haplotype. For example, the term "H-2D^b" means that the D allele being discussed is that which occurs in a mouse strain of the "b" haplotype. Researchers might also identify this allele as D^b and verbalize it as "D of b."

V. Expression of MHC Genes

The expression of MHC genes is tightly and differentially regulated, such that MHC class I is expressed on almost all healthy host cells, but MHC class II expression is limited to APCs. As well, MHC protein expression may be upregulated or induced by cytokines and other stimuli released in a host cell's vicinity. Depending on the type of host cell and the tissue in which it resides, these stimuli may be either constitutively produced or induced during an immune response to injury, pathogens or tumors. For

example, molecules in the walls of invading bacteria stimulate macrophages to produce TNF and LT, and viral infection induces the infected cells to synthesize IFNs. The interaction of these cytokines with specific receptors on a host cell triggers intracellular signaling pathways that activate transcription factors. The activated transcription factors enter the host cell nucleus and bind to 5′ regulatory motifs in the DNA upstream of the MHC genes, altering their expression. An increase in MHC expression facilitates the amplification of an adaptive response by enhancing antigen presentation to T cells.

D. Physiology of the MHC

I. Polymorphism and MHC Restriction

How did the MHC loci come to be so polymorphic? In an ancient, antigenically simple world, a primeval MHC molecule that displayed endogenous and exogenous protein fragments to T cells likely existed but was of very limited (or non-existent) variability. As the world became more antigenically complicated, it was individuals with multiple duplications of this primordial MHC gene that likely survived because they possessed more than one gene dedicated to presenting protein fragments. Perhaps concurrently, different alleles of each gene also evolved, each with a different sequence in the peptide-binding groove. A broader range of peptide binding and presentation molecules would have been generated. Today, the resulting polymorphism at multiple MHC loci ensures that each member of an outbred species is heterozygous at most if not all MHC loci, and thus has a very good chance of possessing at least one MHC allele capable of binding to any given antigenic peptide. For the species as a whole, MHC polymorphism means a large catalog of MHC alleles is spread over the entire population. In the case of a devastating pathogen attack, a significant fraction of the population (but not all individuals) will be able to respond to the pathogen and survive to perpetuate the species. However, the multiplicity of MHC molecules does not allow for a free-for-all in terms of presentation to T cells. A phenomenon called **MHC restriction** exists, which dictates that the epitope seen by a given TCR is a combination of a specific peptide with a specific MHC molecule. The discovery of MHC restriction is outlined in **Box 6-2**.

Box 6-2 The Discovery of MHC Restriction

In the early 1970s, neither the TCR nor the structure it recognized on host cells had been defined. However, in 1974, immunologists Rolf Zinkernagel and Peter Doherty published remarkable results from their studies of T cell responses to lymphocytic choriomeningitis virus (LCMV), a pathogen that is deadly in mice. Their experiments showed that CTLs from virus-infected mice would only kill infected cells derived from inbred strains of mice that had the same MHC haplotype. These researchers hypothesized that CTLs must simultaneously recognize both the antigen from the virus *and* the MHC molecule displaying that antigen on the surface of the target cell. This meant that the daughter effector cells of a naïve Tc cell that initially recognized a given antigen in conjunction with a particular MHC allele would only recognize the same antigen if it was presented in association with exactly the same MHC allele. This phenomenon came to be called "MHC restriction," and finally provided a explanation for why T cells could not bind directly to soluble antigens like B cells can. A great deal of work by many other researchers eventually revealed the nature of the structure that engaged the TCRs of T cells: a small antigen-derived peptide associated with an MHC molecule positioned on the surface of a host cell. In order for a CTL to kill an infected body cell, the TCR of that CTL must have engaged an MHC class I molecule that was displaying peptide from a pathogen antigen and was situated on the surface of the infected cell. Similarly, in order to produce the cytokines needed to support responses by B and Tc cells, the TCR of a Th cell must have engaged an MHC class II molecule that was displaying peptide from a pathogen antigen and was situated on the surface of an APC such as a DC. The fundamental roles of the MHC class I and II molecules were therefore found to be remarkably alike, despite the fact that these molecules were recognized by different T cell subsets. Thus, Zinkernagel and Doherty's revolutionary work helped to establish the very basis of modern cellular immunology. Logical extensions of their work have directly influenced modern vaccine development, transplantation medicine, and the exploration of potential new treatments for autoimmune diseases and cancer. In 1996, these two scientists were justly rewarded for their efforts with the Nobel Prize in Physiology and Medicine.

II. MHC and Immune Responsiveness

Immunologists have long observed that some foreign proteins that provoke strong immune responses in some individuals fail to do so in others. Those individuals failing to mount a response were originally called "non-responders," while those that did react were called "responders." Among responders, there were subtle differences in the level of the response, leading to the description of individuals as either low or high responders. Immunologists soon mapped the genes behind immune responsiveness to the MHC and showed that mice of different MHC haplotypes sometimes respond differently to a given peptide (**Table 6-3**). These variations in response levels to a given antigen can be interpreted as differences in the ability of particular MHC alleles to effectively present peptides from that antigen that can be recognized by T cells. In an inbred population, there is a greater possibility that an individual will be a non-responder; that is, an individual will lack an MHC allele that can lead to specific T cell activation during a challenge with a particular antigen. Two hypotheses, which may not be mutually exclusive, have been proposed to account for non-responsiveness: the *determinant selection* model and the *hole in the T cell repertoire* model.

i) Determinant Selection Model

For a T cell response to be mounted against a foreign protein, a host must possess at least one MHC allele with a groove that accommodates a peptide derived from that protein. Responsiveness then depends on the strength of binding between a given MHC allele and a given determinant (peptide), which in turn depends on structural compatibility. In other words, the MHC proteins in an individual "select" which determinants will be immunogenic in that individual as well as the extent of the response. Since a foreign protein is usually processed into three to four strongly immunogenic peptides, an outbred individual is very likely to possess an MHC allele capable of binding to at least one of these peptides and provoking an immune response. Such an individual is then a responder to this particular antigen, and his/her status as a high or low responder correlates with strong or weak binding of the peptide to MHC, respectively. On the other hand, if none of the individual's MHC molecules can bind to any of the peptides generated from the protein, the individual is a non-responder to this antigen.

The determinant selection model has been supported by experiments in which a given peptide is immunogenic only when it is bound to a particular MHC allele. For example, in inbred mice of the H-2^k haplotype, a certain peptide of an influenza virus protein readily provokes an immune response. More specifically, the determinant is recognized when presented to T cells on the MHC class I H-$2K^k$ molecule. However, this same peptide fails to stimulate T cells in mice of the H-2^b haplotype. Instead, a peptide from a different part of the same influenza virus protein triggers an antiviral response in H-2^b mice when presented on the MHC class I H-$2D^b$ molecule.

TABLE 6-3 MHC Haplotype Correlated with Immune Responsiveness

Mouse Strain	H-2 Haplotype	Response to TGAL* Peptide
C3H	k	Low
C3H.SW	b	High
A	a	Low
A.BY	b	High
B10	b	High
B10BR	k	Low

*TGAL, synthetic peptide containing lysine, alanine, tyrosine and glutamic acid residues.

ii) Hole in the T Cell Repertoire Model

Immune non-responsiveness may also result from tolerance mechanisms. It may be that, in non-responders, a particular foreign peptide–MHC combination very closely resembles the structure of a self peptide–MHC combination, such that any T cell clones capable of recognizing the foreign peptide–MHC combination were eliminated as autoreactive during the establishment of central tolerance. In a non-responder, this would result in a missing T cell specificity or a "hole" in the T cell repertoire relative to the repertoire of a responder.

III. MHC and Disease Predisposition

An individual's MHC haplotype determines his/her responsiveness to immunogens. If an individual cannot mount an appropriate immune response to an immunogen associated with infection or cancer, the individual will likely suffer disease. If an immune response is mounted when it is inappropriate, disease in the form of autoimmunity or hypersensitivity (including allergy) can result. The direct link between immune responsiveness and particular MHC alleles means that certain MHC haplotypes may predispose individuals to specific susceptibilities or disorders.

NOTE: The Human Genome Project (HGP), which identified the ~25,000 genes in human DNA and sequenced the 3 billion base pairs of the human genome, was completed in 2003. The HGP has made it possible to determine the genomic locations and DNA sequences of HLA alleles with unprecedented precision. By comparing the sequences of HLA alleles derived from diverse populations around the world, researchers have examined the links between certain HLA alleles and disease susceptibility among people of many different nationalities. In addition, relationships between numerous non-HLA genes and disease incidence have been discovered. Studies that utilize the vast information of the HGP to map the location of genes linked to particular diseases have been termed "Genome-Wide Association Studies" (GWAS).

In humans, many of the disorders linked to the possession of specific MHC alleles manifest as **autoimmune disease**. Autoimmune disease results when self-reactive T cell clones escape the tolerance mechanisms that would normally prevent these cells from entering or acting in the periphery. The individual may then possess T cells that can recognize self components and may attack tissues expressing these components. For example, type 1 (insulin-dependent) diabetes mellitus is thought to arise from an autoimmune attack on antigens expressed by the insulin-producing β cells of the pancreatic islets. Immune destruction of the β islet cells results in insulin deficiency and thus diabetes. For unknown reasons, the HLA-DQ8 allele is eight times more prevalent in groups of humans suffering from type 1 diabetes than it is in healthy populations. Similarly, 90% of Caucasian patients suffering from a degenerative disease of the spine called ankylosing spondylitis carry the HLA-B27 allele, whereas only 9% of healthy Caucasians do. Additional autoimmune diseases and their association with particular HLA alleles are included in **Table 6-4**. Note, however, that mere possession of a

TABLE 6-4 Examples of HLA-Associated Disorders in Humans

Disease	Examples of Associated HLA Alleles
Ankylosing spondylitis	B27
Birdshot retinopathy	A29
Celiac disease	DR3, DR5, DR7
Graves' disease	DR3
Narcolepsy	DR2
Multiple sclerosis	DR2
Rheumatoid arthritis	DR4
Type 1 diabetes mellitus	DQ8, DQ2, DR3, DR4

predisposing HLA allele is not usually sufficient to cause disease; other genetic and environmental factors are thought to be involved. A complete discussion of autoimmune disease is presented in Chapter 19.

This concludes our discussion of the structure and physiology of the MHC. In the next chapter, titled "Antigen Processing and Presentation," we describe the derivation of antigenic peptides and how MHC molecules associate with these peptides and present them to T cells.

Chapter 6 Take-Home Message

- The MHC class I and II genes in the MHC encode cell surface proteins that present peptides to T cells.
- Non-classical MHC class I and class II genes as well as MHC class III genes also constitute part of the MHC.
- MHC class I and class II proteins are heterodimeric molecules with highly variant N-terminal domains that form a peptide-binding site.
- MHC class I is expressed on almost all cells in the body and generally presents peptides of endogenous origin.
- MHC class II is expressed by APCs and generally presents peptides of exogenous origin.
- MHC class I interacts with the CD8 coreceptor found on Tc cells and CTLs, while MHC class II interacts with the CD4 coreceptor on naïve and effector Th cells.
- The MHC genes are characterized by polygenicity, extreme polymorphism and codominant expression.
- Outbred populations are heterozygous at the MHC loci, while inbred mice are homozygous. Syngeneic individuals have the same MHC genotype, while allogeneic individuals have different MHC genotypes.
- Differences in MHC alleles are largely responsible for transplant rejection and variations in immune responsiveness to a given pathogen.
- Expression of particular MHC alleles is linked to autoimmune disease predisposition.

Did You Get It? A Self-Test Quiz

Section A

1) Describe the derivation of the term "major histocompatibility complex."
2) What polypeptide chains make up the MHC class I molecule? MHC class II?
3) What cell types express MHC class I? MHC class II?
4) To what type of cells does MHC class I present peptide? MHC class II?
5) Name five regions of the HLA complex and describe the nature of their gene products.
6) Name five regions of the H-2 complex and describe the nature of their gene products.
7) Name three non-classical MHC genes and describe the nature of their gene products.
8) What MHC-like proteins present non-peptidic antigens to T lineage cells?

Section B

1) Why do MHC proteins have relatively low binding affinity for peptides?
2) What is an anchor residue?
3) Do self peptides bound to MHC usually provoke immune responses? If not, why not?
4) What domains form the peptide-binding grooves of MHC class I molecules? Of MHC class II molecules? How do these grooves differ in structure?
5) Describe two ways in which the peptides binding to MHC class I differ from those binding to MHC class II.

Section C

1) Can you define these terms? polygenicity, monomorphic, polymorphic, allele, codominance
2) Which MHC molecule is more common in a human: DQA/DRB or DQA/DQB?
3) Where is amino acid variation concentrated in the MHC protein and why?
4) Distinguish between the terms "heterozygous" and "allogeneic."
5) How many different types of HLA-DQ proteins are likely present in an outbred individual?
6) Can you define these terms? haplotype, ancestral haplotype

Did You Get It? A Self-Test Quiz—Continued

7) What does the term "H-2K^b" represent? How would you verbalize this term?

8) What effect does inflammation have on MHC expression?

Section D

1) Why does inbreeding sometimes put a species at risk for decimation by a pathogen?

2) What is MHC restriction, and how was it discovered?

3) Outline two theories accounting for variation in immune responsiveness to an antigen.

4) What is a GWAS, and how might it be helpful for studying HLA alleles?

Can You Extrapolate? Some Conceptual Questions

1) The HLA-A2 MHC class I molecule is found to present an 8 amino acid peptide "X" to a T cell clone. In the lab, researchers produce two mutated versions of X, which they call "Xm1" and "Xm2." Xm1 is mutated at amino acid #4 of the peptide sequence, and Xm2 is mutated at amino acid #7. Each peptide is tested to see if its presentation by HLA-A2 will stimulate the T cell clone. How would you explain the following observations?
   a) The T cell activation response to Xm1 is less than the T cell activation response to X.
   b) The response to Xm1 is greater than the response to X.
   c) The response to Xm2 is less than the response to X.
   d) The response to Xm2 is greater than the response to X.

2) MHC molecules and Ig molecules can both be said to show diversity. Briefly explain how the origin of diversity in each of these cases is fundamentally different.

3) You are a researcher studying the amino acid sequences of MHC class I and class II alleles. In which domains of these molecules would you expect to find the most allelic variations? Explain briefly.

Would You Like To Read More?

Apostolopoulos, V., Yuriev, E., Lazoura, E., Yu, M., & Ramsland, P. A. (2008). MHC and MHC-like molecules: Structural perspectives on the design of molecular vaccines. *Human Vaccines*, *4*(6), 400–409.

Corse, E., Gottschalk, R. A., & Allison, J. P. (2011). Strength of TCR-peptide/MHC interactions and *in vivo* T cell responses. *Journal of Immunology*, *186*(9), 5039–5045.

Hofstetter, A. R., Sullivan, L. C., Lukacher, A. E., & Brooks, A. G. (2011). Diverse roles of non-diverse molecules: MHC class Ib molecules in host defense and control of autoimmunity. *Current Opinion in Immunology*, *23*(1), 104–110.

Liao, W. W., & Arthur, J. W. (2011). Predicting peptide binding to major histocompatibility complex molecules. *Autoimmunity Reviews*, *10*(8), 469–473.

Marrack, P., Scott-Browne, J. P., Dai, S., Gapin, L., & Kappler, J. W. (2008). Evolutionarily conserved amino acids that control TCR-MHC interaction. *Annual Review of Immunology*, *26*, 171–203.

Menier, C., Rouas-Freiss, N., Favier, B., LeMaoult, J., Moreau, P., & Carosella, E. D. (2010). Recent advances on the non-classical major histocompatibility complex class I HLA-G molecule. *Tissue Antigens*, *75*(3), 201–206.