Histocompatibility Antigens

Secondary article

Narinder K Mehra, *All India Institute of Medical Sciences, New Delhi, India*

Article Contents

The histocompatibility antigens are cell surface glycoproteins expressed on nucleated cells whose major function is to bind peptides within the cell and present them at the cell surface for inspection by T cells of the immune system. They are a part of the major histocompatibility complex (MHC) that exists in most vertebrate species and whose antigens were originally defined as the most important molecules involved in the rejection of transplanted tissues, exchanged between individuals of the same species.

Major Histocompatibility Complex

The human major histocompatibility complex (MHC) gene cluster, referred to as the human leucocyte antigen (HLA) system, spans a region of about 4000 kilobases (kb) (4×10^6 nucleotides) on the short arm of chromosome 6 in the distal portion of the 21.3 band (**Figure 1**). The murine MHC, termed the H-2 complex, is located on chromosome 17. Data from physical mapping, deoxyribonucleic acid (DNA) cloning and sequencing of the MHC region from a number of laboratories have revealed the presence of a large number of genes with variable expression and function. Several different types of gene are arranged in the form of three regions: class I, class II and class III. Most of these genes are polymorphic, arranged close together and are generally inherited *en bloc* as a haplotype.

Human HLA-A, B and C loci resemble mouse H-2 K, D and L, and are referred to as class I MHC molecules, whereas human HLA-DR, DQ and DP resemble mouse I-A and I-E, and are called class II MHC molecules. Similar polymorphic loci have been found in every vertebrate species examined. While the class I molecules are expressed on most nucleated cells and platelets, class II expression is limited to selected immunoreactive cells: B lymphocytes, macrophages, dendritic cells and immune activated T cells. Under some circumstances, MHC class II expression can be transiently induced on other cells of the body by endogenous factors, such as interferon γ. Class III genes (central genes) placed between class I and class II genes code for factors involved in the complement system and in various nonimmune functions.

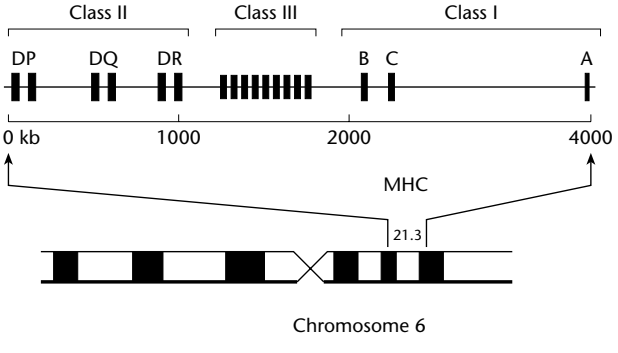

Figure 1 Schematic representation of the human leucocyte antigen (HLA) complex on chromosome 6. The extent of the class I, class II and class III regions is marked and the major genes are marked by bars. MHC, major histocompatibility complex.

MHC class I genes

The class I region is the most telomeric part of the MHC complex. Although 36 genes have been defined so far in this region, HLA-A, B and C are the most important since their products have been well defined as 'classical transplantation antigens'. They are characterized by high degree of polymorphism in most vertebrate species. Other human class I genes that show sequence homology to classical loci include HLA-E, F, G, H, and a set of five MIC (MHC class I related) genes (MIC A–E). These have reduced expression, restricted to certain tissues such as thymus, liver, intestine or placenta, and low polymorphism. Of the five MIC genes, only MIC A and MIC B, situated between the tumour necrosis factor (TNF) and HLA-B locus, are expressed. Closely related to these genes lies the haemochromatosis disease candidate gene, designated *Hfe*.

HLA-A, B and C molecules are heterodimeric glycoproteins consisting of a MHC-encoded α or heavy chain of about 44 kDa and a nonMHC-encoded light chain (β_2-microglobulin) of 12 kDa molecular weight (**Figure 2**). The α chain is some 350 amino acid residues long and can be divided into three functional regions: external, transmembrane and intracytoplasmic. The extracellular portion of the heavy chain is folded into three globular domains, α1, α2 and α3, each of which contains stretches of about 90 amino acids encoded by separate exons. While the α1 and α2 domains take part in antigen binding (antigen-binding domains), the α3 domain is essentially conserved. The transmembrane region (23–25 amino acid residues) spans the lipid bilayer of the plasma membrane while the cytoplasmic region (30–32 amino acid residues) has elements of the cytoskeleton.

The β_2-microglobulin is a soluble protein encoded by a gene located on chromosome 15 in humans; it associates noncovalently with the heavy α chain. Cells lacking β_2-

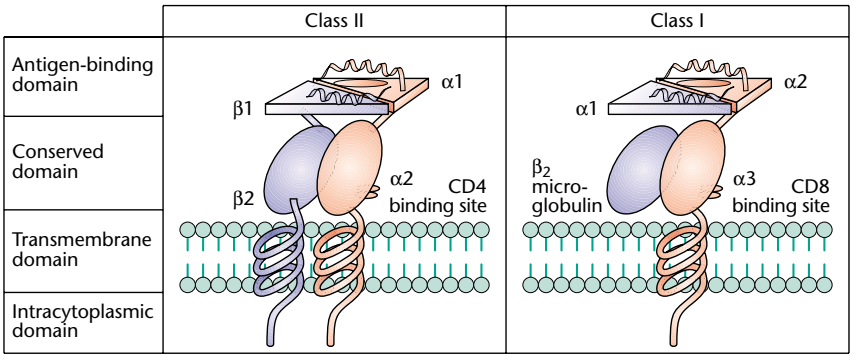

Figure 2 Biochemical structure and three-dimensional folding of human leucocyte antigen (HLA) class I and II molecules. The two antigen-binding domains in class I molecules are contributed by the heavy α chain, whereas both α and β chains contribute the same in the class II molecule.

microglobulin are deficient in the expression of MHC class I molecules, indicating its important role in the transport of these molecules to the cell surface. Recently, X-ray crystallography of the HLA-A2 molecule has provided an exciting leap forward in our understanding of the three-dimensional structure of the MHC (Bjorkman *et al.*, 1987a). It has become clear that:

1. the membrane-proximal structures (α3 and β_2-microglobulin) are folded to form immunoglobulin-like domains, and thus MHC molecules are considered to be a part of the immunoglobulin gene superfamily. The α3 and β_2-microglobulin are responsible for imparting general shape to the class I molecule by providing a platform for top-lying polymorphic domains. Further, α3 domain contains binding sites for the α chain of the CD8 glycoprotein, which is important for the recognition of antigen by cytotoxic T cells.
2. the region distal from the membrane is formed by α1 and α2 domains; each consisting of four β strands forming a platform which constitutes the antigen-binding site. Thus the two α helixes form the sides of a cleft whose floor is formed by a plane of eight antiparallel β-pleated sheets. The dimension of this cleft ($2.5 \times 1.0 \times 1.1$ nm) is large enough to accommodate peptides ranging from 10 to 15 amino acids in length, depending on how the peptide is folded (Bjorkman *et al.*, 1987b).

MHC class II genes

MHC class II region extends over 1000–1200 kb with at least six subregions, termed DR, DQ, DP, DO, DN and DM. Structurally, the class II molecules are similar to class I molecules and are expressed as heterodimers on the cell surface with one heavy α chain (molecular weight 34 kDa) and one β chain (molecular weight 29 kDa) of integral membrane glycoproteins (**Figure 2**). Three-dimensional structural differences between the two include an altered position of the immunoglobulin-like β2 domain relative to that of the α3 domain of class I HLA, and considerable changes in the peptide-binding site.

The DR region contains multiple, highly polymorphic β genes and only one invariant α gene. The conventional serologically defined DR molecules (DR1–DR18) are coded for by the *DRB1* gene, whereas the DR52 and DR53 specificities are encoded by the *DRB3* and *DRB4* genes respectively. *DRB2*, *DRB6*, *DRB7*, *DRB8* and *DRB9* are pseudogenes without a first domain exon. The DQ subregion contains five genes, *DQA1*, *DQA2*, *DQB1*, *DQB2* and *DQB3*, of which *DQA2*, *DQB2* and *DQB3* are not known to be expressed. In contrast, both *DQA1* and *DQB1* are functional and polymorphic, expressing four different types of DQ molecules by different 'cis' and 'trans' combinatorial events. The DP subregion contains two α and two β genes, with *DPA2* and *DPB2* being pseudogenes. *DPB1* shows extensive polymorphism, while *DPA1* displays limited polymorphism. DO, DM and DN lie between the DQ and DP loci, and have very limited polymorphism, if any.

In addition to the above, two groups of nonHLA genes have been identified in the MHC class II region. The first group is ABC transporter genes called *TAP1* (transporter associated with antigen processing or transporter of antigen peptides) and *TAP2* genes (Spies *et al.*, 1989). *TAP1* and *TAP2* gene products associate as a heterodimer that is involved in the transport of antigen fragments produced in the cytoplasm, into the lumen of the endoplasmic reticulum. The second group is proteasome-related genes that includes *LMP2* (low molecular mass polypeptide or large multifunctional protease) and *LMP7* genes (Monaco and McDevitt, 1986). The products of *LMP2* and *LMP7* genes are large cytoplasmic proteolytic complex molecules that contain multiple catalytic sites. LMP complex is involved in the production of multiple peptides simultaneously from the same substrate to produce peptides better suited for MHC class I binding.

MHC class III genes (central genes)

The central region of MHC has no structural or functional correlation with the class I or class II region. Presently, at least 39 genes have been located in a 680-kb stretch of DNA within this region. This includes genes encoding proteins involved in the immune system: the complement genes C4, C2 and Bf (factor B), the TNFα and TNFβ (lymphotoxin) genes and the heat-shock protein (HSP70) genes. Genes with no obvious association with the immune system have also been identified in this region. These include G7a (valyl transfer ribonucleic acid synthetase). Further, two B-cell associated transcript genes, *BAT2* (G2) and *BAT3* (G3) are novel genes in the class III region that encode large proline-rich proteins with molecular masses of 228 and 110 kDa, respectively, while the RD gene encodes a 42-kDa intracellular protein.

Apart from the complement components, this region also contains two genes coding for the steroid hormone, 21-hydroxylase or CYP21 genes. The 21-hydroxylase genes associate very closely with the C4A and C4B genes. Of these, the 21B gene (*CYP21B* gene) is more functional, and deficiency of this leads to the clinical condition of congenital adrenal hyperplasia or salt-wasting disease. The *CYP21A* gene on the other hand, is most often deleted in certain specific HLA haplotypes, particularly the extended haplotype HLA-A1, B8, DR3, SCO1 in Western caucasians. Such a haplotype is known to be associated with insulin-dependent (type 1) diabetes mellitus.

Polymorphism and linkage disequilibrium

One striking feature of the products of the classical class I and II loci is their extreme degree of polymorphism. Indeed, no other loci are known to have a similar degree of polymorphism, which means that an exceptional inter-individual variability exists as far as the HLA profile of a population is concerned. With the introduction of DNA-based methods of HLA typing, extensive molecular polymorphism has been discovered in each of the relevant HLA class I and class II loci, which is far in excess of the polymorphism obtained by serological methods. For example, currently at least 288 alleles have been described for HLA-B, 139 for HLA-A, 87 for HLA-C, 259 for HLA-DR, 65 for HLA-DQ, and 107 for HLA-DP (website: http://imgt.cnusc.fr.8104). HLA antigens being diploid in nature, an individual inherits two alleles of each of the above loci, one each from either parent. Family studies have shown that recombination in the HLA region is rare (less than 1%), and thus a complete set of alleles on the same chromosome is usually inherited as a haplotype. The two HLA haplotypes in an individual derived after family testing constitute the genotype, whereas the total HLA antigen profile is the phenotype. Theoretically, siblings in a family have a 25% chance of being HLA identical, a 50% chance of being HLA haploidentical (sharing one parental haplotype only), and a 25% chance of being HLA unidentical.

The high degree of polymorphism in the HLA region is a result of several factors. These relate to the major biological function of the HLA molecules, which is to bind peptides. Environmental influence can lead to the presence of specific disease-inducing peptides that can bind variable DRB molecules. Theoretically, several million genotypic combinations (approximately 150 billion or even more) are possible in the HLA system. According to Klein (1987), such a polymorphism is not only advantageous for an individual, but even more for the survival of the species surrounded by many different and often changing pathogens. Additionally, racial admixture can lead to the formation of new recombinants as a result of exchange of genetic material between alleles of the same locus or due to point mutation and other genetic events (Mehra *et al*., 1996). Thus the polymorphism is probably essential for efficient functioning of the system.

A close study of the different populations has shown that certain combinations of HLA alleles occur more often than would be expected on the basis of their individual gene frequencies. This nonrandom association of the alleles of two HLA loci found together on the same HLA haplotype is termed linkage disequilibrium and is expressed in terms of delta. It is found to vary among different populations. Although the reason for these associations is unknown, linkage disequilibrium may be the consequence of natural selection for or against a specific gene combination, or it may be due to the fact that the population has not yet reached equilibrium. The stable relationship between alleles has been taken to suggest that these combinations represent preserved ancient or ancestral haplotypes.

Most people are heterozygous for each of the HLA loci due to the extensive polymorphism at each of them. However, sometimes both parents may have one and the same allele on one of their HLA haplotypes. This is the most common in populations with a high degree of consanguinity and in alleles with a relatively high frequency in the population. The child inheriting such a haplotype from both parents is homozygous for that allele. For the same reason, an individual can be homozygous for a complete haplotype, due to a relatively high frequency of some haplotypes in a given population. Biologically, a heterozygote might have an advantage in being able to present a wider array of pathogen-derived peptides than a homozygote, given the fact that different MHC allelic products bind different arrays of peptides.

Minor Histocompatibility Antigens

The minor histocompatibility (mH) antigens are defined as nonMHC-encoded cell surface processed peptides which, in association with MHC, contribute to graft rejection,

usually less severe than that due to MHC disparity. Incompatibility of mH antigens between donor and recipient constitutes a potential risk for graft-versus-host disease (GVHD) or graft failure after bone marrow transplantation both from HLA phenotypically matched unrelated donors as well as from HLA genotypically identical sibling donors. Compared with polymorphic MHC molecules, mH antigens can evoke only MHC-restricted cytotoxic T lymphocyte (CTL) and proliferative T helper cell responses, but not antibody-mediated immunity. Human mH-specific T-cell responses have also been described after organ transplantation, blood transfusions and pregnancy.

Tissue distribution

Using congenic strains, more than 40 different mH loci have been defined in the mouse. The number of mH antigens in humans is not yet fully known. Because of their function as targets of GVHD following bone marrow transplantation, estimates vary from less than a dozen to more than 1000. In addition to the autosomes, the mH antigens in humans show sex-linked expression. The male-specific mH antigen, H-Y is encoded by a gene present on the Y-chromosome and functions to control spermatogenesis. The nonsex-linked mH antigens represent a group of five antigens, HA-1–HA-5, that occur with a variable frequency in the healthy population, from less frequent (7–16%) to most frequent (69–95%). Their segregation pattern in families follows a Mendelian mode of inheritance, independent of HLA. Tissue distribution of mH antigens reveals differential expression: while some (H-Y, HA-3 and HA-4) have a broad tissue distribution, being expressed on haematopoietic cells as well as on cells of epithelial origin, others (HA-1 and HA-2) are limited to cells of the haematopoietic lineage only. Besides, all mH antigens are present on both clonogenic leukaemic precursor cells as well as circulating leukaemic cells of lymphocyte and myeloid origin.

Clinical applications

The molecular nature of mH antigens has been defined by using immunochemical isolation and purification procedures following sensitization of target cells for T-cell recognition. Results of these studies indicate that mH antigens are peptides from polymorphic self proteins. They are derived from evolutionarily conserved genes with important biological functions. In the clinical context, certain combinations of mH antigens and MHC alleles are likely to be more antigenic than others. Identification of which minor antigens are immunodominant in association with defined MHC alleles offers the chance to type for these polymorphisms at the DNA level in donor–recipient pairs for risk assessment, choice of immunosuppression and possible peptide immunomodulatory treatment. As more antigens are defined by means of molecular techniques, identification of 'major minors' (functionally more important mH antigens) from 'minor minors' will aid in immunomodulatory approaches.

It is now clear that mH antigens play an important role not only in development of GVHD but also for preventing rejection by inducing tolerance. The H-Y peptide information can be used for prenatal diagnosis in sex-linked congenital abnormalities and for investigating the minimal residual disease and chimaerism. Other areas in which mH antigens appear to have a pivotal role include: (i) immunotherapy of leukaemias using CTLs specific for mH antigen peptide and (ii) corneal allograft rejection whereby the mH antigens act as major barriers while the MHC incompatibility plays a minor role.

Biological Significance

Although only about one-thousandth of the total human genome, the HLA system contains several genes that have important functions in biology and medicine. Besides their major role in donor selection for organ and bone marrow transplantation, other important areas in which HLA has provided great help include paternity determination, identification of susceptibility or predisposing genes for a wide variety of diseases, particularly those with infectious and autoimmune aetiology, prediction of 'risk' development for disease in families, anthropological characterization of different races and ethnic groups, and for understanding the control and regulation of the immune system.

The MHC was originally defined as the set of most important molecules involved in the rejection of transplanted tissues exchanged between members of the same species. Currently data from large transplant organizations have shown conclusively that better HLA matching between donor and the recipient, coupled with nonspecific immunosuppression, offers the best way to achieve long-term graft survival following not only renal transplantation but also that of the heart. Data collected at the University of California at Los Angeles Transplant Registry in the USA, involving several thousand kidney transplants carried out worldwide, indicate that patients transplanted with two haplotype matched kidneys have graft half-life of 25 years compared with 12 years for one-haplotype matches and 6.5 years for poorly matched cadaveric kidneys (Terasaki, 1991). For bone marrow transplantation, complete HLA identity of the donor and the recipient is essential for obtaining optimum results.

The major biological function of the MHC is to bind peptides for presentation to T cells. Following discovery in the early 1970s of the immune response genes and their mapping within the class II region of the mouse H-2 system

(Benacerraf and McDevitt, 1972), the cellular recognition of foreign antigens was explained through elegant experiments performed by Zinkernagel and Doherty (1974). Their studies revealed that cytotoxic T cells could recognize a virus-infected target cell only in the context of class I molecules, a phenomenon called MHC restriction. For their pioneering discovery about the T-cell antigen recognition system, and the biological function of the MHC, Rolf Zinkernagel and Peter Doherty received the Nobel Prize in Medicine and Physiology for the year 1996.

Peptide–MHC interactions

Immune surveillance helps to protect us from foreign intruders such as bacteria and viruses. The first step in the process is specific recognition, allowing distinction to be made between self and nonself. The high degree of MHC polymorphism is advantageous for increasing the range of responses that different individuals can mount. More specifically, highly diverse binding repertoires would prevent escaping of immune recognition by mutation of a few crucial antigenic peptides. The HLA molecules bind all peptide fragments that may fit into the peptide-binding groove of class I and II molecules, without distinction of whether they are derived from self or nonself proteins. The function of the peptide-presenting HLA molecule is illustrated schematically in **Figure 3**. MHC class I ligands originate from endogenous sources, mainly from self proteins, or viral proteins that infect a cell, or cytosolic proteins released by bacteria or protozoan parasites. These peptides are delivered to CD8 + CTLs for elimination by the endogenous processing pathway. In contrast, the MHC class II ligands are generated by the degradation of proteins from the exogenous sources for presentation to CD4 + helper T cells. Such peptides are derived from cell surface proteins, soluble proteins, or proteins from a virus, bacterium or protozoan parasite that invades or is phagocytosed by the cell.

The processing pathway in both situations utilizes highly specialized machinery, which works most effectively. The endogenous proteins in the cytoplasm of cells are digested into 8–10 amino acid peptides by the LMP complex, and these gain access to the endoplasmic reticulum (ER) via the TAP transporter and are finally exported to the cell surface along with the self class I MHC molecule. The CD8 + CTLs see this MHC–peptide complex through the T-cell receptor (TCR) and other coreceptors, leading to killing of the target cell. In the class II pathway, antigens taken from outside the cell by endocytosis are digested by the proteolytic enzymes in the endosomal compartment into small peptide fragments. The MHC class II molecule with its α and β chains assembles in the ER with the invariant chain (Ii), to form a stable trimolecular complex which inhibits the binding of endogenous or self peptides in the ER. The trimolecular complex is efficiently transported out of the ER and is targeted to a post-Golgi compartment in the peripheral cytoplasm. The Ii is subsequently cleaved in the endoplasm, opening up the cleft for peptide occupancy. The MHC–peptide complex thus formed is recognized by the CD4 + helper T lymphocytes.

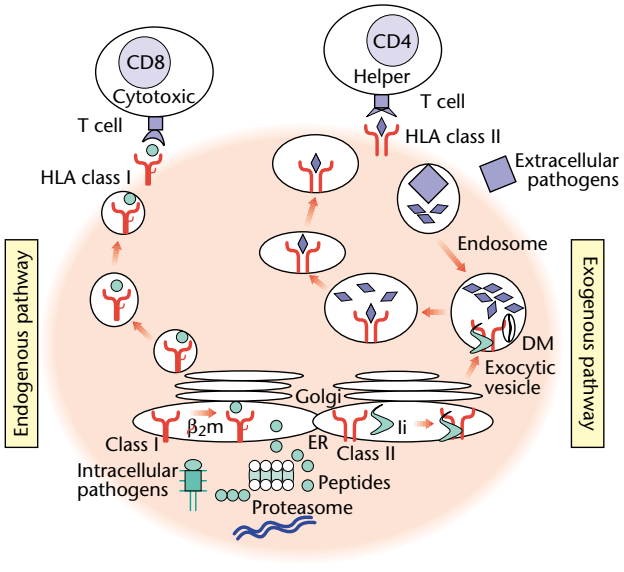

Figure 3 Presentation of peptide fragments to T cells by human leucocyte antigen (HLA) class I and II molecules. β_2m, β_2-microglobulin; ER, endoplasmic reticulum; DM, a heterodimer encoded by the *DMA* and *DMB* genes; Ii, invariant chain. See text for further details.

Role of MHC molecules

The functional divergence of MHC class I and II molecules is perhaps necessary to increase the efficiency of the immune system for handling a broad spectrum of peptides. Although the general rules for peptide–MHC interactions for both classes of MHC molecules are essentially similar, X-ray crystallographic studies have indicated that distinct peptides bind to class I or class II molecules owing to differences in their binding clefts. In accordance with the distribution of hydrogen bonds in the peptide-binding groove of the MHC, the anchor residues are placed at the terminal ends of the class I groove. Contrarily, the binding forces are distributed throughout the class II groove and ensure bonds between the peptide's backbone and the class II molecule. Since the cleft of class I MHC molecules has closed ends, peptides bound by them are short, generally 9–10 amino acids. The residues occupy deep pockets (particularly pockets B and F) in the antigen-binding site. On the other hand, the MHC class II cleft is more open ended, accommodating peptides that are more variable in size and much longer than nonamers, generally 14–18 amino acids (Stern *et al.*, 1994). The groove-anchored

stretch of the peptide (core sequence) is exactly nine amino acids long, with its flanking parts protruding out of the groove's ends (Stern *et al*., 1994). In relation to the MHC class II motif structure, relative positions of five peptide-binding pockets designated as P1, P4, P6, P7 and P9 have been defined, each of which accommodates specific peptide side-chains.

It is clear that, while extracellular surveillance is taken care of by B cells and antigen-presenting cells (e.g. macrophages, dendritic cells, fibroblasts, Langerhans cells, Kupffer cells), intracellular surveillance is subserved mainly by T cells. The TCR recognizes peptides only if they are presented in association with the self-MHC molecule: CD8 + T cells are restricted by MHC class I–peptide complexes and CD4 + T cells by class II–peptide complexes. Since CD4 T cells are placed at the initial phase of the immune response and provide helper signals to both CD8 + T cells as well as B cells, they are central to both extracellular and intracellular surveillance.

References

Benacerraf B and McDevitt HO (1972) Histocompatibility linked immune response genes. *Science* **175**: 273–279.

Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL and Wiley DC (1987a) Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* **329**: 506–512.

Bjorkman PJ, Saper MA, Samraoni B, Bennett WS, Strominger JL and Wiley DC (1987b) The foreign antigen binding site and T-cell recognition region of class I histocompatibility antigens. *Nature* **329**: 512–518.

Klein J (1987) Origin of major histocompatibility complex polymorphism. The transpecies hypothesis. *Human Immunology* **19**: 155–162.

Mehra NK, Rajalingam R and Giphart MJ (1996) Generation of DR51 associated DQA1, DQB1 haplotypes in Asian Indians. *Tissue Typing* **47**: 85–89.

Monaco JJ and McDevitt HO (1986) The LMP antigens: a stable MHC-controlled multisubunit protein system. *Biochemistry* **15**: 416–426.

Spies T, Blank G and Bresnahan M (1989) A new cluster of genes within the human major histocompatibility complex. *Science* **243**: 214–217.

Stern LJ, Brown JH, Jardetzky TS *et al*. (1994) Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature* **368**: 215–221.

Terasaki PI (1991) Histocompatibility testing in transplantation. *Archives of Pathology and Laboratory Medicine* **115**: 250–254.

Zinkernagel RM and Doherty PC (1974) Restriction of *in vitro* T-cell mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. *Nature* **248**: 701–702.

Further Reading

Abbas AK, Lichtman AK and Pober JS (1997) *Cellular and Molecular Immunology*. Philadelphia: WB Saunders.

Bodmer JG, Marsh SGE, Albert ED *et al*. (1997) Nomenclature for factors of the HLA system, 1996. *Tissue Antigens* **49**: 297–321.

Charron D (1997) *Genetic Diversity of HLA: Functional and Medical Implications*, vols I and II. Paris: EDK Publishers.

McCluskey J and Peh CA (1999) The human leucocyte antigens and clinical medicine: an overview. *Reviews in Immunogenetics* **1**: 3–20.

Simpson E, Roopenian D and Goulmy E (1998) Much ado about minor histocompatibility antigens. *Immunology Today* **19**: 108–112.

Terasaki PI (1990) *History of HLA: ten recollections*. Los Angeles, CA: UCLA Tissue Typing Laboratory.