# What is the MHC?

Stephen H. Powis and Daniel E. Geraghty

*The major histocompatibility complex (MHC) has been studied for many years, but it still continues to puzzle geneticists and immunologists alike. Nevertheless, the study of MHC evolution is providing insights into some of its enigmas, and these were discussed at a recent meeting*.*

In the fifty years since the rules of histocompatibility were first established, the major histocompatibility complex (MHC) has become one of the most intensively studied regions of the vertebrate genome. Therefore, it seems rather odd that one of the first questions posed at a recent workshop on MHC evolution was 'what exactly is the MHC?' Nevertheless, the question was astute because it encompasses a variety of fascinating conundrums. For example, why are nearly all the genes encoding class I and II molecules so closely linked and what mechanisms have kept them together? What about notable exceptions such as genes encoding CD1 and $\beta_2$-microglobulin ($\beta_2$-m) – should we consider them fully fledged MHC members, or outriders that have set out alone? When and why did the *TAP* and *LMP* genes become part of the MHC? Furthermore, what of the class III region – did it barge in on the complex by chance or does it contain genes that are important for MHC integrity? Indeed, do any of the nonimmunological genes within the complex have a role in MHC function? How and why is MHC polymorphism generated, and at what rate?

## The MHC through the ages

Polymorphic genes encoding MHC molecules have been found as far back as cartilaginous fish, but attempts to isolate MHC genes from cyclostomes, the most primitive vertebrates, have so far failed (M. Kasahara, Sapporo). MHC class I and class II genes have been found linked in all well-characterized vertebrates, and thus show strong synteny. The single exception among vertebrates is the zebrafish (*Brachydanio rerio*), but it is not known whether this is an isolated case or a general rule for fish (J. Klein, Tübingen). Interestingly, some syntenic conservation does exist between zebrafish and mouse class I genes. The zebrafish class I genes map 19 cM from the *Brachydanio no-tail* locus, and the homologue in the mouse [the *Brachury (T)* locus] maps 14 cM centromeric of *H-2K*.

Genes in the MHC class III region, such as those encoding complement component 2 (C2), C4 and heat shock protein 70 (Hsp70), have been found linked to class I and II genes in several species, including *Xenopus*. A *C4*-like exon has recently been identified near the class I and II genes of the chicken (C. Auffray, Villejuif), a species in which the class III region was previously thought to be separated from the rest of the MHC. The chicken is unusual because it contains a second cluster of MHC class I and II genes, *Rfp-Y*, which appears to segregate independently of the main class I and II loci. However, recent evidence has shown that the two clusters are on the same 12–16 Mb microchromosome, but separated by an extremely high frequency of recombination (M. Miller, Duarte). The reason for this separation is not known, but may represent functional specialization.

## Sex, parasites and the MHC

MHC class I and class II molecules are highly polymorphic, but what drives allelic diversity? One of the simplest explanations is that pathogen-driven selection is largely, if not solely, responsible, but evidence to support this has been difficult to ascertain.

A. Hill (Oxford) demonstrated malaria-driven selection at *HLA-B*, and has recently studied the association between severe *Plasmodium falciparum* malaria and HLA class II alleles. In East Africa, *HLA-DR*0101* confers resistance to severe malaria, whereas in West Africa, *HLA-DRB1*1302* confers resistance. This difference may be due to allelic variation in the HLA-DR-binding epitopes derived from the major strains of *P. falciparum*. Thus, polymorphism in parasite epitopes may underlie the geographical variation in HLA-DR association.

The interaction between parasite and host is likely to be more complex than previously thought (Hill; S. Gilbert, Oxford). Statistical analysis suggests that certain *P. falciparum* isolates that have allelic differences in an HLA-B35-specific cytotoxic T lymphocyte (CTL) epitope are preferentially found together in infected individuals. This may be advantageous to the parasite because the CTL epitope from one isolate antagonizes the predominant CTL response against the other isolate. W. Potts (Gainesville) provided an intriguing insight into the relationship between MHC molecules and pathogens through the observation of MHC class I-deficient ($\beta_2$-m-knockout) mice in semi-natural colonies. Somewhat surprisingly, these animals showed relatively undiminished health and vigour. Nevertheless, selection was operating against them through reduced reproductive success of $\beta_2$-m$^-$ territorial males and reduced weaning success of $\beta_2$-m$^-$ females. However, it was not clear whether reduced expression of the placental Fc receptor, which is known to associate with $\beta_2$-m, might affect litter size.

*The 4th International Workshop on MHC Evolution was held in St Augustine, FL, USA, on 4–8 March 1995.

Although polymorphism at the peptide-binding site of MHC molecules may be critical for protection against infection, pathogen-driven selection is not necessarily the only mechanism driving allelic diversity. It has been known for some time that mating-type selection in mice favours MHC heterozygosity, and that the genetic factors underlying this selection map to the MHC. However, the molecular basis for the phenomena has been difficult to explain. One possible answer may come from the observations of D. Geraghty and M. Janer (Seattle), who have identified olfactory receptor-like genes within the human and mouse class I regions. Unlike olfactory-receptor genes on other chromosomes, the MHC-resident genes are considerably more polymorphic, a quality required of genes directly involved in mating-type selection. In a study of a well-defined population of Hutterites, a reproductively isolated group of European migrants found in several locations in the northern USA and Canada, C. Ober (Chicago) found that mating patterns were not culturally defined but could be explained by genetic factors linked to the MHC. The olfactory-receptor-like genes found in the MHC are highly conserved between human and mouse, both in their primary sequence and in the extent of allelic polymorphism, raising the fascinating possibility that the mating patterns of Hutterites may be related to olfactory-receptor-like genes in the human MHC. Has the MHC hijacked the olfactory system as a supplementary means of maintaining heterozygosity?

### Rate and mechanism of MHC evolution

Are MHC genes evolving slowly, with conservation of alleles between species, or more rapidly, with new variation being generated in response to changing pathogenic stimuli? New evidence supports both these views. Trans-species MHC polymorphism was reported in the historical population constituted by Darwin's finches (Klein). Within the 13 species of finches found on the Galapagos Islands, some class II alleles were unique to individual species, while others were common among different species. Using a mutation rate derived from other studies, these observations were used to estimate a founding-population size of several thousand. However, an analysis of *HLA-DPB1* showed that the rate of MHC mutation, at least at this locus, may be considerably faster than some models have assumed (H. Erlich and G. Zangerberg, Almeda). Although there are 65 *DPB1* alleles, only two to three differences exist at any one amino acid position, resulting in very localized, discrete polymorphic sequence motifs. This suggests that gene conversion may play a significant role in the generation of new alleles. To test this, the *DPB1* locus in sperm was analysed in a well-controlled series of experiments that distinguished authentic gene-conversion events from novel alleles created by artefacts of the polymerase chain reaction. These studies suggested a mutation rate that is two orders of magnitude higher than that found at non-MHC loci.

While the debate over rates of mutation at the antigen-binding site of MHC molecules continues, some investigators have turned their attention to MHC evolution at other sites. Studying introns 1 and 2 of *HLA-DRB1*, U. Gyllensten (Uppsala) found almost as much variation in introns as exons. Thus, intron variation predicts exon variation and identical exon-alleles on different haplotypes have identical intron-alleles. Diversity in HLA class I promoters was analysed by A. Vallejo (Rochester), who found that the genetic distances between *HLA-A* promoters is greater than that between *HLA-B* promoters. This may be related to the observation that the interferon-stimulated response element (ISRE) of *HLA-B* is functional while the *HLA-A* ISRE is not. In an evolutionary study of self-peptides bound by HLA class I molecules, A. Hughes (Pennsylvania) observed that these peptides were generally derived from the conserved hydrophilic regions of highly conserved hydrophobic proteins. Thus, the mechanisms by which peptides are selected for presentation by class I molecules may favour the conserved regions of proteins. In the case of parasite-derived peptides, such a mechanism would be advantageous to the host if it reduced the likelihood of escape mutants. Intracellular antigen processing may also exert selective pressures on MHC alleles that recognize particular peptide motifs. P. Parham (Stanford) noted that an unusually high number of *HLA-B* alleles preferentially bind peptides with proline at position 2. It was postulated that such peptides may be more common because the proteasome cannot cleave at proline residues.

### Specialized MHC genes

One of the most intriguing observations of recent years is that MHC genes have been adapted for specialized tasks throughout evolution. In some cases, adaptation has been conserved between species, but in other instances, species have remained unique. The nonclassical class I genes are a good example, and have been studied to the greatest extent in the mouse. To date, 58 mouse nonclassical class I genes, pseudogenes and gene fragments have been identified, of which the most distal is *H-2M3*. This molecule preferentially binds *N*-formylated peptides, which makes it particularly suited to the presentation of prokaryotic peptides. The basis for this unusual specificity has been elucidated by X-ray crystallography (K. Fischer Lindahl, Dallas). At 2.1 Å resolution, H-2M3 resembles a class Ia molecule, but the peptide-binding groove differs such that the bound peptide is shifted along one residue. As a result, the A pocket is closed, the P1 side-chain lodges in the B pocket and the His9 residue becomes the critical residue for binding the formyl group.

Surprisingly, species other than rodents (including humans) do not appear to have evolved H-2M3-like molecules, and have thus missed out on a seemingly useful anti-bacterial mechanism. By contrast, CD1, which is encoded outside the MHC, has evolved a different antimicrobial strategy that may be conserved between species (M. Kronenberg, Los Angeles). Human CD1b binds a mycobacterial glycolipid and thus may have a role in specific infections such as tuberculosis. Using a phage-display library, mouse CD1 was shown to bind peptides with a sequence motif comprising bulky,

hydrophobic amino acids. Peptide length did not appear to be critical. The natural ligand for mouse CD1 is not known, but may well also be lipid.

Class I and II genes arose by duplication, and most of these have remained closely linked throughout evolution, suggesting that a specific mechanism is keeping them together. Nevertheless, it is still possible that this linkage simply represents chance, and that there has not been enough evolutionary time for the genes to become completely separated. However, this cannot be the case for the *TAP* and *LMP* genes, whose sequences are unrelated to each other, and to the genes encoding class I and II molecules. The most convincing explanation for the presence of the *TAP* genes within the MHC is found in the rat, where *TAP2* proteins encoded by different alleles have been shown to transport different peptides into the endoplasmic reticulum. The rat RT1.A$^a$ class I molecule has a highly charged C-terminal pocket that predominantly binds peptides with a C-terminal arginine, and thus requires the rat *TAP2A* allele, which is capable of supplying such peptides. The rat *TAP2B* allele has a more selective specificity for peptide transport and cannot supply these peptides. On the basis of these observations, it has been proposed that the association of *TAP2A* and *RT1.A$^a$* on the same haplotype confers a selective advantage.

G. Butcher (Cambridge, UK) reported further work in support of the above hypothesis. Sequence analysis of a number of rat *RT1* haplotypes demonstrated that alleles encoding TAP2A were predominantly linked to those encoding RT1.A predicted to bind peptides with a C-terminal arginine, whereas alleles encoding TAP2B were predominantly linked to those encoding RT1.A unable to bind arginine. The functional polymorphism of rat TAP2 is associated with a complex polymorphism involving 25 amino acids. Using site-directed mutagenesis, five residues located in two distinct regions of the TAP2 hydrophobic domain were identified as the amino acids responsible for this difference (J. Howard, Cologne; Butcher). TAP polymorphism has also been analysed in humans, mice and, most recently, in gorillas (D. Lawlor, Houston), but no evidence has yet been found in these species for the complex polymorphism that exists in the rat. If this is the mechanism keeping the *TAP* genes within the MHC, it will be important to identify other species that exhibit functional TAP polymorphism. Similarly, it will be interesting to determine whether any species has functional LMP polymorphism.

## Conclusion

In summary, what is the MHC? For most species, the simplest answer is perhaps the original definition: it is the locus (or region of the genome) that encodes the cell-surface proteins primarily responsible for allograft rejection. Nevertheless, with specialized nonclassical MHC genes, antigen-presenting genes, olfactory-receptor genes and a host of other genes of unknown function, we now know that the MHC is one of the most complex and informative regions of the vertebrate genome. Before the end of the century, the MHCs of several species will have been completely sequenced and many outstanding questions should have been answered. One cannot help but wonder whether a few more unexpected secrets will be revealed before that task is completed.

We thank John Trowsdale and Adrian Hill for reviewing the manuscript. S.H.P. is an MRC Clinician Scientist Fellow.

*Stephen H. Powis is at the Clinical Science Laboratories, United Medical and Dental Schools of Guy's and St Thomas's Hospital, London, UK SE1 9RT; Daniel E. Geraghty is at the Human Immunogenetics Program, Fred Hutchinson Cancer Research Centre, Seattle, WA 98104 USA.*