HLA and infection

Human leukocyte antigens (HLA) have been the subject of intensive research because of their importance in organ transplantation and the finding of many HLA associations with autoimmune disorders. Although their central role in the control of immune responsiveness has long been appreciated, it is only recently that the importance of HLA variation in determining susceptibility to infectious diseases has been confirmed. Indeed it is now increasingly accepted that associations of particular alleles with fatal infectious diseases have led to the evolution of the extraordinary diversity of HLA types that are observed in modern-day populations. I shall review here some of the approaches used to define associations between HLA and infectious diseases, and outline how these promise to be useful in increasing our understanding of protective immunity against various infectious pathogens, and may eventually allow more rational vaccine design.

HLA molecules and genes

Human leukocyte antigens are cell surface glycoproteins expressed in the majority of tissues. They come in two classes (Fig. 1). HLA class I antigens (HLA-A, -B and -C), composed of a polymorphic heavy chain and an invariant light chain, β_2 microglobulin, present peptides (of 8 or 9 amino acids) derived from the breakdown of cytoplasmic proteins to CD8 positive, often cytotoxic, T lymphocytes [1, 2]. HLA class II antigens (HLA-DR, -DQ and -DP) comprise α and β chains, both of which may be polymorphic, and present peptides (of perhaps 13–17 amino acids in length [3]) to CD4 positive T lymphocytes which secrete various regulatory cytokines that stimulate macrophages and provide 'help' for antibody formation [4]. Peptides presented by class II molecules usually derive from internalised exogenous antigens. Crystallographic analysis of some HLA molecules has provided an informative picture of their three-dimensional structure [5]. A peptide binding groove, formed of two α helices and a base of a β-pleated sheet, allows peptides of a certain length and binding affinity to be presented to a T lymphocyte. The amino acids which vary between different HLA types cluster around this peptide binding groove so that different HLA molecules can bind and present a different set of peptides.

The HLA genes form part of the major histocompatibility complex (MHC) on chromosome 6. The arrangement of these highly polymorphic genes and a recent estimate of the number of alleles at each locus is shown in Fig. 2 [6]. This extensive coding region diversity is unequalled anywhere else in the genome and the possible reasons for this extraordinary diversity of MHC genes, seen in humans and most other vertebrates studied, have generated considerable debate amongst evolutionary geneticists. The distribution of the sequence variation provides one clue: there are far more amino acid changing substitutions in nucleotides which code for residues forming part of the peptide binding groove. This implies that natural selection must have been acting on this binding site to maintain a diversity of structures which will bind different peptides [7].

There is evidence of specialisation within the class I and class II families. HLA-C is the most weakly expressed class I locus and HLA-C restricted T cells are rare. HLA-B is more polymorphic than HLA-A and might be functionally more important. HLA-DRB1 is the most polymorphic and most highly expressed class II locus and DR-restricted T lymphocytes are the most readily isolated. Interestingly, however, many autoimmune disease associations may be primarily with DQ alleles and it has been speculated that this may reflect a particularly important role for DQ molecules in shaping the repertoire of T cell receptor molecules during lymphocyte development in the thymus [8].

HLA typing and disease associations

HLA types were first identified serologically, and for almost two decades this remained the only means of distinguishing allelic types. A series of international workshops ensured sufficient standardisation of these reagents so that HLA typing by serological methods became widely used: in renal transplantation matching, in thousands of disease association studies, and in anthropological comparisons of human populations. The diseases for which associations could be confirmed were often 'autoimmune' in nature, with few other clues available as to pathogenetic processes, thus providing a spur to ever greater dissection of the HLA associations. Although a few clearly infectious diseases were looked at in case-control studies, and possible associations reported, overall the results were inconsistent, and in any case the aetiology of infectious conditions was felt to be understood.

Then came the revolution in molecular biology leading to the cloning and sequencing of all the HLA genes and many of their alleles. Rapid molecular typing methods, initially using restriction fragment length polymorphisms and then amplification by the polymerase chain reaction, have confirmed what was long suspected—that serological reagents give a very incom-

ADRIAN V. S. HILL, BM, DPhil, MRCP, *Wellcome Senior Research Fellow in Clinical Science, John Radcliffe Hospital, Oxford.*
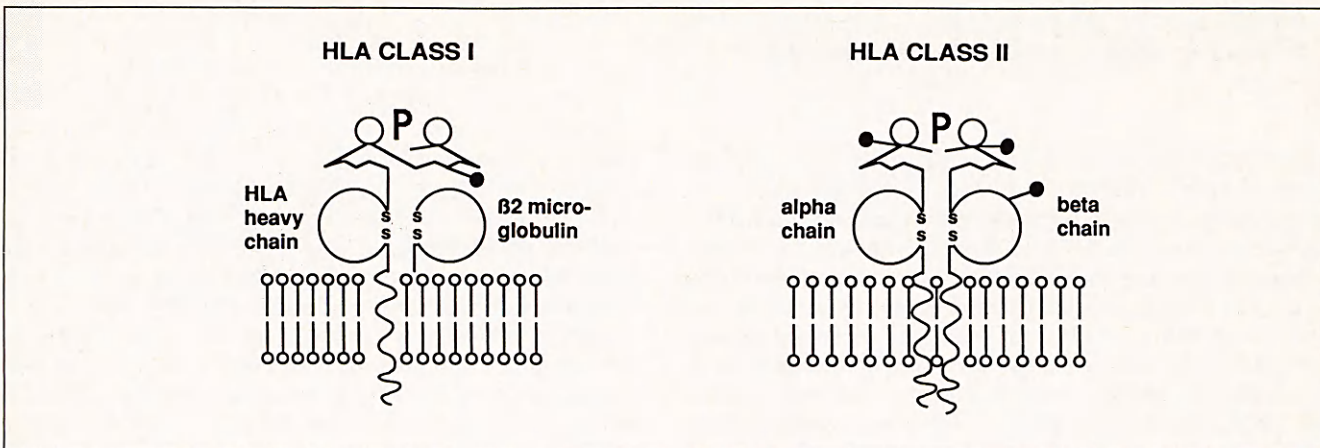
Fig. 1. *Diagrams of the structure of HLA class I and class II molecules.* Class I molecules are composed of a highly polymorphic heavy chain and the smaller invariant β2 microglobulin. Both the α and β chains of HLA class II molecules are very variable in the domains which comprise the membrane-distal peptide binding groove. The position of a peptide in the groove, made up of sides of α helices and a floor of β pleated strands, is indicated here by the letter P.

plete picture of the extent of polymorphism at most HLA loci. The improved molecular resolution, together with studies of diverse ethnic groups, has permitted the localisation of certain associations to a particular HLA locus, eg HLA-DRB1 for rheumatoid arthritis and mainly HLA-DQ for juvenile diabetes, or to particular residues within a locus [9]. Other advantages of molecular techniques, particularly for studies of paediatric diseases or tropical populations, are that only tiny amounts of blood are needed, and the requirement for fresh, or liquid nitrogen stored, lymphocytes is removed.
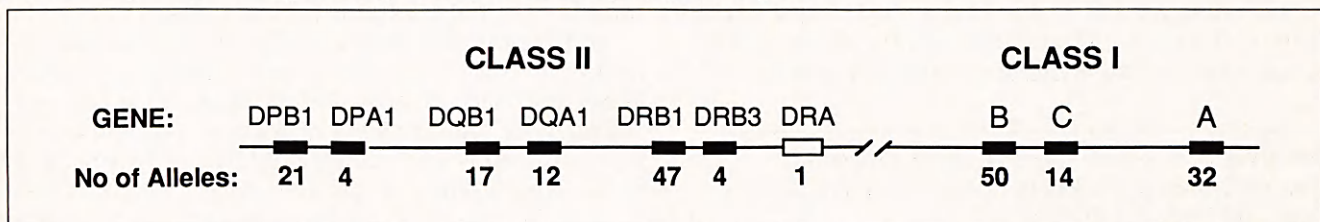
However, this progress has still failed to yield the hoped for insights into the mechanisms of autoimmunity, which appear always to lie just around the corner. Nonetheless, the considerable experience gained in the design, execution, analysis and interpretation of HLA case-control studies has been valuable in looking again at the neglected area of HLA and major infectious diseases. Additionally, those who study HLA associations with several infections now have what autoimmune disease researchers lack—good candidates for the immunodominant antigens which may interact with HLA molecules to produce any observed associations.

HLA and infection: approaches

A survey of the literature on HLA associations with infectious diseases makes difficult, and often confusing, reading. In addition to the changing HLA nomenclature, there seems to be little consistency in the associations reported by different workers. One problem here is statistical. In most HLA studies associations are sought between disease and a multiplicity of HLA types so that one association reaching the conventional 5% level of statistical significance would be expected for every 20 alleles typed. This accounts for some of the false-positive associations in the early literature. More recently it has become accepted that one should multiply significance levels (p values) by the number of comparisons performed. However, some ambiguity

Fig. 2. *Map of the arrangement of polymorphic HLA genes in the human major histocompatibility complex on chromosome 6.* A minimum estimate of the number of alleles at each locus is shown below the genes. The class I genes encode polymorphic heavy chains for HLA-A, -B and -C. The HLA class II molecule HLA-DQ has two polymorphic chains, α and β, which are specified by the DQA1 and DQB1 genes, respectively. The α chain of HLA-DR is invariant but the DRB1 gene encoding the major DR β chain is highly polymorphic.

	CLASS II								CLASS I		
GENE:	DPB1	DPA1	DQB1	DQA1	DRB1	DRB3	DRA		B	C	A
No of Alleles:	21	4	17	12	47	4	1		50	14	32

remains: should one correct for all alleles or the number of alleles per locus; were several disease categories assessed; need one correct for extremely rare alleles? A good way around this difficulty is to do a study in two halves, identifying a possible association in the first half and re-testing this single association in the second.

After allowing for multiple comparisons, statistically significant associations have been reported for certain infectious diseases, notably leprosy and HIV infection (Table 1). However, often the associations differ between studies. Assuming that the trivial explanation of poor matching of cases and controls does not hold, there are other explanations for this inconsistency. An old idea was that different associations in particular ethnic groups might reflect different linkages of the marker HLA serotype to a common disease susceptibility gene, situated near to it on chromosome 6. This now seems unlikely since it has become clear that HLA genes actually are immune response genes rather than closely linked markers. A more attractive proposal derives from new molecular information on the extent of polymorphism in major infectious pathogens. Striking diversity has been observed in immunodominant antigens of, for example, malaria parasites and HIV, with some evidence of geographical variation. This molecular diversity in infectious agents may account for variable HLA associations, although this remains to be demonstrated.

Perhaps the greatest problem in assessing HLA associations with infection has been the small size of most studies, so that associations are often unconvincing. This reflects a genuine difficulty in accumulating large series of cases of acute life-threatening infections caused by a single pathogen in most countries with advanced HLA typing facilities. Contrast this with the ease of studying patients with chronic autoimmune diseases, who often obligingly cluster in specialist clinics. This inherent selection bias is also reflected in the large number of HLA studies of leprosy. Molecular typing offers a solution to some of these difficulties by facilitating the logistics of studies in developing countries, reducing the time and costs involved in large scale typing, and improving the detection of non-caucasian alleles.

HLA and malaria

As an example of how molecular typing of HLA antigens in an infectious disease can yield interesting and potentially useful results, I will outline a study of genetic susceptibility to malaria conducted recently with several colleagues from The Gambia and Oxford [10]. To maximise the power of the study to detect significant effects of HLA variation, children with life-threatening malaria were studied. In Africa, severe malaria in children presents as almost always either cerebral malarial or severe malarial anaemia. The former syndrome is the commoner in The Gambia, the latter in coastal

Table 1. A selection of some recent HLA associations with major infectious diseases.

Disease	Population	Antigen	Frequency
Leprosy	Korean	DR2	Increased [13]
Malaria	Gambian	Bw53, DRw13.02	Decreased [10]
AIDS	British	A1-B8-DR3	Increased [15]
AIDS	American	A1-B8-DR3	Increased [16]

Kenya [11]. These severe malaria cases were compared with several control groups. The primary controls were 500 children without malaria who were carefully matched for age (mean <3 years) and area of residence around the Gambian capital, Banjul. The feasibility of HLA typing these control children using as little as 400 microlitres of blood greatly facilitated collection of these samples. Additional control groups included children with uncomplicated malaria and healthy adults so that, in all, over 2,000 Gambians were studied.

A two-stage approach was taken in analysing HLA class I antigens. After serologically typing about half of the 600 cases of severe malaria, a single apparent association was seen: the frequency of the HLA-Bw53 antigen was reduced amongst the children with each of the two forms of severe malaria. To confirm this possible association the remaining samples were tested just for HLA-Bw53, using a new method based on the polymerase chain reaction. This second half of the study showed a similar significant protective effect with HLA-Bw53, confirming the suggested association.

Analysis of HLA class II haplotypes, composed of HLA-DR and -DQ antigens, revealed another protective association, with a haplotype bearing the DRw13.02 antigen (ie the 02 subtype of DRw13). In this case the association was greater (and only statistically significant) for cases of severe malarial anaemia. Several groups had shown that the HLA-Bw53 antigen is common only in sub-Saharan Africans, where it is found at frequencies of up to 40%, and the protective HLA class II haplotype is also relatively African-specific. So, as with the geographical distribution of the thalassaemias and some other haemoglobinopathies, the prevalence of these HLA types associated with malaria resistance in Africans suggests that natural selection by *P. falciparum* has contributed to their present high frequencies. However, it remains to be determined whether these associations are also found in other parts of Africa where different strains of *P. falciparum* may predominate.

The magnitude of these protective effects can be compared to the protection against severe malaria afforded to carriers of sickle haemoglobin. In the same Gambian study these carriers had a ten-fold reduction in risk of developing severe malaria compared with non-carriers. Children with the protective HLA alleles were at approximately half the risk of

developing severe disease (Table 2). Thus the sickle haemoglobinopathy is clearly the most useful variant to have. However, in public health terms, or from the point of view of an evolutionary geneticist, the HLA effects might be considered to be at least as important. This is because close to half the population of children in The Gambia have one of these HLA types, whereas only 13% are sickle haemoglobin carriers. Hence, it can be calculated that the HLA alleles are responsible for preventing at least as many cases of severe malaria as HbS does. In making these simple extrapolations, it is important to remember that other HLA alleles, found at such low frequencies that they cannot be adequately assessed in case-control studies of reasonable size, may also have significant effects on malaria susceptibility. Also, by affecting susceptibility to other infections, HLA alleles that protect against malaria might have unpredictable effects on overall survival. Nonetheless, the Gambian study suggests that, if sufficient cases can be studied, important associations with several other infections may be found by employing detailed molecular HLA typing.

Leprosy

Amongst infectious diseases, leprosy has probably been the most frequently studied in the search for HLA associations. These studies have been reviewed by Blackwell [12]. Although the overall impression is one of inconsistency, particularly in terms of HLA class I associations, more recent studies have frequently found an association between susceptibility and the HLA-DR2 antigen, particularly in Asian populations [eg 13]. This association appears stronger for tuberculoid than for lepromatous leprosy, and it has been suggested that the latter may be primarily associated with HLA-DQw1. Our current recognition of a multiplicity of DQw1 subtypes raises the possibility that a stronger association may exist with one of these. However, studies of the segregation of HLA haplotypes in multiple case families have led to the view that HLA antigens may be important in determining the type of leprosy an individual develops rather than influencing susceptibility to leprosy *per se*. This would appear compatible with the quite different immune responses seen in tuberculoid and lepromatous leprosy, and would explain why the markedly greater disease concordance for monozygotic than dizygotic twins is not reflected in a clear HLA association. Furthermore, recent complex segregation analyses of multigenerational pedigrees [14] point to a major role for a single non-HLA-linked gene in determining susceptibility to leprosy *per se*.

AIDS

Many studies of HLA and various manifestations of HIV infection have now been reported. For a disease that afflicts so many people, the sample sizes in many

Table 2. Carrier rates (for HLA types, antigen frequencies) of some genetic variants associated with resistance to severe malaria. The estimate of % protection to an individual is from reference 10, with the HLA class II haplotype value given for protection against severe malarial anaemia.

	Carrier rate		Protection
	North Europe	The Gambia	
Haemoglobin AS	0%	13%	92%
HLA-DRB1*1302-DQB1*0501	<1%	20%	55%
HLA-Bw53	1%	25%	41%

of these surveys have been disappointingly small. This is reflected in a variety of apparent associations that have not been confirmed by other studies. However, other associations, found on more than one occasion, are of particular interest. The common caucasian haplotype HLA-A1-B8-DR3 was associated with an increased probability of seroconversion and more rapid disease progression in a series of haemophiliacs [15]. This was supported by the finding of the same haplotype associated with rapid decline in CD4+ lymphocyte counts in a cohort of seropositive homosexuals [16]. Although not all studies have found this association, its plausibility is enhanced by the well documented association between this haplotype and several autoimmune diseases. Whatever increased immunoreactivity predisposes to autoimmune disease might also affect the rate of decline in CD4+ lymphocyte counts. Other groups have reported that the HLA class I allele, B35, may also be associated with more advanced disease [17]. Although studies of a hundred or so cases continue to appear, it seems likely that far larger studies may be required to show very convincing associations.

Infections cause HLA polymorphism

The unparalleled diversity of HLA genes, and their homologues in the major histocompatibility complex of other species, has given rise to a large number of theories to explain why this extreme polymorphism exists. A few years ago, with the analyses of MHC genes in other higher primates, it became clear that the polymorphism of HLA genes is very old and certainly predates our speciation. The human HLA-A1 gene, for example, is far more akin to a particular chimpanzee allele than it is to any other human class I allele. This gave rise to the suggestion that all HLA types might be effectively equivalent in terms of survival, or selectively neutral [18]. However, it is now clear that to maintain this range of HLA types over millions of years some type of balancing selection needs to be in operation: certain forces tending to increase the frequencies of particular HLA types, other selection pressures acting

to decrease their frequency. Without this most alleles seen today would have been lost through genetic drift. The question has been: what is the nature of this selection?

One remarkable type of selection has recently been found to operate in mice. Potts and colleagues [19] observed that mice (who can smell MHC differences) will mate preferentially with partners differing in MHC type so that offspring tend to be heterozygous and a diverse array of alleles is kept in the population. This disassortive mating seems unlikely to occur in humans but various other theories have been advocated (Table 3).

The proposal that infectious pathogens are the major cause of HLA diversity has been particularly attractive because defence against microorganisms is a primary function of HLA molecules. By being heterozygous rather than homozygous for HLA alleles, an individual may respond to a greater range of foreign peptides in an infectious pathogen, and is thereby more likely to be able to mount a protective immune response against it. The argument against this theory has been the lack of documented HLA associations with infectious diseases in humans and almost all other species. The HLA associations in the Gambian malaria study [10] now suggest that small effects on susceptibility to infection might be fairly common but that large studies, employing detailed typing methods, may be needed to document these convincingly. Although such effects are less striking than associations with many autoimmune diseases (in which there may be a breakdown in tolerance to just a single non-polymorphic antigen), they are large enough to retain a considerable diversity of HLA types. Interestingly, the fluctuating intensities of many infectious diseases may also play a role in maintaining HLA diversity by altering selection pressures on particular alleles as epidemics come and go.

What use are HLA associations with infection?

If HLA associations are so difficult to identify, is it worth knowing about them? Probably not, in terms of measuring individual susceptibility to most infections. Rather, their value lies in what they may be able to tell us about the nature of protective immunity to infectious diseases. I will use the example of malaria again, but this is just one of several major infections in which we have little knowledge of the mechanisms of naturally acquired protective immunity. The finding of an HLA class I association with protection from severe malaria suggests that, as in mouse models of malaria [20], HLA class I restricted T lymphocytes, presumably cytotoxic T cells, play a role in natural immunity. Identification of the antigenic targets of such cells is of major interest in vaccine development. Similarly, the finding of HLA associations in an infection where the parasite must present an enormous number of epi-

Table 3. Some of the theories that have been proposed to account for the marked degree of polymorphism of major histocompatibility complex genes. The first two are discussed in the text; 'parasite' embraces all infectious microorganisms.

Theories of MHC polymorphism
Parasite driven selection
Disassortive mating
Maternal–fetal incompatibility
High mutation rate
Ancient neutral polymorphism

There is little evidence in humans that HLA dissimilarity between mother and fetus improves fetal survival rates. Unusually high mutation rates have not been observed and, although HLA polymorphism is ancient, population genetic theory argues against its neutrality.

topes to the host immune system says something about the nature of protective immune responses. Because all these potential T cell epitopes will have a great variety of HLA restriction patterns, and yet a particular HLA association with decreased susceptibility is detectable, it seems likely that, amongst the large number of immune responses that the parasite elicits, only a very few are usefully protective. The protective HLA alleles may be a key to identifying which parasite antigens are protective, by correlating HLA types with immune responses to candidate parasite antigens.

However, potentially the most direct approach to exploiting an HLA association with an infection has become possible by very recent advances in the analysis of the peptides presented by HLA molecules. It has been known for some years that HLA molecules present peptides derived from the breakdown of internalised or cytoplasmic antigens; but it has only recently become possible to elute sufficient peptides from purified HLA molecules to analyse their length and amino acid sequences. Although such studies are still few, it has been found that particular class I molecules will only bind peptides consisting of 8 or 9 amino acids, with a sequence motif characteristic of the particular HLA type [2, 5]. This requirement for a particular type of peptide in each different HLA molecule's groove seems to relate to allelic variation in so-called pockets within the groove, which must bind particular amino acid side chains. It has also been possible to identify slightly longer peptides from HLA class II molecules [3]. These techniques open up the possibility not only of rapidly identifying peptides within proteins that might be epitopes for particular HLA alleles, but also of characterising foreign peptides bound to HLA molecules of infected cells. This, together with some other techniques of identifying peptide epitopes, should facilitate analysis of the mechanisms of HLA associations and identify epitopes for inclusion in future multivalent subunit vaccines.

References

1 Townsend A, Bodmer H. Antigen recognition by class I-restricted T lymphocytes. *Ann Rev Immunol* 1989;**7**:601–24.

2 Falk K, Roetzschke O, Stevanovic S, Jung G, Rammensee H-G. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* 1991;**351**:290–6.

3 Rudensky AY, Preston-Hurlburt P, Hong S-C, Barlow A, Janeway CA. Sequence analysis of peptides bound to MHC class II molecules. *Nature* 1991;**353**:622–7.

4 Schwartz RA. T-lymphocyte recognition of antigen in association with gene products of the major histocompatibility complex. *Ann Rev Immunol* 1985;**3**:237–61.

5 Madden DR, Gorga JC, Strominger JL, Wiley DC. The structure of HLA-B27 reveals nonamer self-peptides bound in an extended conformation. *Nature* 1991;**353**:321–5.

6 Bodmer J, Marsh SGE, Albert ED, *et al.* Nomenclature for factors of the HLA system, 1990. *Tissue Antigens* 1991;**37**:97–104.

7 Hughes AL, Nei M. Pattern of nucleotide substitution at major histocompatibility class I loci reveals overdominant selection. *Nature* 1988;**335**:167–70.

8 Altmann DM, Sansom D, Marsh SGE. What is the basis for HLA-DQ associations with autoimmune disease? *Immunol Today* 1991;**12**:267–70.

9 Bell JI, Todd JA, McDevitt HO. The molecular basis of HLA disease association. *Adv Hum Genet* 1989;**18**:1–4.

10 Hill AVS, Allsopp CEM, Kwiatkowski D, *et al.* Common West African HLA antigens are associated with protection from severe malaria. *Nature* 1991;**352**:495–500.

11 Greenwood BM, Marsh K, Snow RW. Why do some African children develop severe malaria? *Parasitol Today* 1991;**7**:277–81.

12 Blackwell JM. Bacterial infections: In: *Genetics of resistance to bacterial and parasitic infections*, D Wakelin, JM Blackwell, eds. London: Taylor and Francis, 1988:63–101.

13 Kim SJ, Choi IH, Dahlberg S, *et al.* HLA and leprosy in Koreans. *Tissue Antigens* 1987;**29**:146–53.

14 Abel L, Demenais F. Detection of major genes for susceptibility to leprosy and its subtypes in a Caribbean island: Desirade Island. *Am J Hum Genet* 1988;**42**:256–66.

15 Steel CM, Ludlam CA, Beatson D, *et al.* HLA haplotype A1 B8 DR3 as a risk factor for HIV-related disease. *Lancet* 1988;**i**:1185–8.

16 Kaslow RA, Duquesnoy R, Van Raden M, *et al.* A1, CW7, B8, DR3 HLA antigen combination associated with rapid decline of T-helper lymphocytes in HIV-1 infection. *Lancet* 1990;**335**:927–30.

17 Scorza Smeraldi R, Fabio G, Lazzarin A, *et al.* HLA-associated susceptibility to acquired immunodeficiency syndrome in Italian patients with HIV infection. *Lancet* 1986;**ii**:1187–9.

18 Klein J. Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. *Hum Immunol* 1987;**19**:155–62.

19 Potts WK, Manning CJ, Wakelund EK. Mating patterns in semi-natural populations of mice influenced by MHC genotype. *Nature* 1991;**352**:619–21.

20 Romero P, Maryanski JL, Corradin G, *et al.* Cloned cytotoxic T cells recognise an epitope in the circumsporozoite protein and protect against malaria. *Nature* 1989;**341**:323–5.

Address for correspondence: Dr A. V. S. Hill, Molecular Immunology Group, Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DU.