Genes of the Major Histocompatibility Complex in Mouse and Man

Michael Steinmetz and Leroy Hood

Summary. The genes of the major histocompatibility complex code for cell-surface molecules that play an important role in the generation of the immune response. These genes and molecules have been studied intensively over the last five decades by geneticists, biochemists, and immunologists, but only recently has the isolation of the genes by molecular biologists facilitated their precise characterization. Many surprising findings have been made concerning their structure, multiplicity, organization, function, and evolution.

In the past 3 years fundamental new insights into the structure, organization, and evolution of the genes of the major histocompatibility complex (MHC) in mouse and man have resulted from the cloning and characterization of these genes. The molecules encoded by the MHC genes play a critical role in the rejection of organ transplants and the control of the immune response. The molecular mechanisms of these functions are being studied by means of DNA-mediated gene transfer into cells and embryos.

Major Histocompatibility Complex in Mouse and Man

Three families of genes. The MHC appears to be present in all vertebrates. Because there are allelic forms of the genes of the MHC it has been possible to produce alloantisera that identify the gene products of three families of MHC genes in mouse and man (*1*). These three families are denoted class I, class II, and class III. The class I and class II molecules, typified by transplantation antigens and the Ia (I-region associated) antigens, respectively, are integral membrane proteins involved in the recognition reactions that permit the immune system to distinguish between self and nonself. The class III family encodes several components in the activation pathway of complement and will not be discussed further.

Serological analyses of recombinant MHC chromosomes in mice and man have been used to construct genetic maps for the murine and human MHC's (Fig. 1). The MHC, also known as the H-2 complex, is located on chromosome 17 and spans about 2 centimorgans of DNA which corresponds to approximately 4000 kilobase pairs (kbp) of DNA. The complex is divided into six regions called K, I, S, D, Qa, and Tla. The classical H-2 complex comprises the genes of the K through the D regions and encodes the class I transplantation antigens K, D, and L, the class II and the class III molecules. At least three molecules, denoted Qa-1, Qa-2, and TL, which are structurally closely related to the class I transplantation antigens are encoded by genes located to the right of the H-2 complex (*2*). The two categories of class I molecules are distinguished by their tissue distributions in that the transplantation antigens are present on virtually all somatic cells of the mouse, whereas the Qa antigens are preferentially expressed on B cells and T cells and the TL antigen on thymocytes and certain leukemia cells. The two categories are furthermore distinguished by the fact that the genes encoding transplantation antigens are extremely polymorphic with more than 50 different alleles present at the K and at the D loci, whereas the Qa and TL antigens are much less polymorphic. In addition, alleles at the K and D loci are very different from each other, whereas Tla alleles appear to be much more homologous (*3*). Individual alleles are denoted by a capital letter for the gene and a superscript for the H-2 genotype (also called haplotype) of the inbred strain, for example, the K^d and D^d genes of the BALB/c mouse. Two distinct types of Ia molecules, I-A and I-E, have been identified. The Ia molecules have limited tissue distribution—primarily on B cells, macrophages, and T cells. The Ia molecules also exhibit extensive serological polymorphisms.

The human MHC, also referred to as the HLA complex, is contained on chromosome 6 and encompasses about 3 centimorgans of DNA extending perhaps over 6000 kbp of DNA. The human class I genes differ from their mouse counterparts in genetic organization in that the A, B, and C genes are contiguous to one another, whereas in the mouse the K gene is separated from the other class I genes by the class II and class III gene families (Fig. 1). Possible human homologs to the mouse Qa and TL antigens have been characterized serologically and biochemically but their genes have not yet been mapped (*4*). Three distinct types of human class II molecules have been described which are designated DR, DC, and SB (*5*).

Class I and class II molecules. Structures of the class I and class II molecules are similar (Fig. 2) (*6*, *7*). The class I polypeptide contains three external domains, each about 90 residues in length, a transmembrane region, and a cytoplasmic domain. The third external domain is noncovalently associated with β_2-microglobulin, a small polypeptide that shows homology to the constant region domains of immunoglobulins and is not encoded in the MHC. The class II molecules are composed of two noncovalently associated polypeptide chains, denoted α and β, both of which are encoded in the MHC. Each of these polypeptides has two external domains that are of similar size as for class I molecules, a transmembrane region, and a small cytoplasmic domain.

Functions of class I and class II molecules. Transplantation antigens restrict the recognition of foreign antigens by cytotoxic T cells (*8*). For example, when cells of mouse or man are infected with a virus, foreign viral antigens often are expressed on the cell surface. In order for cytotoxic T cells to destroy these infected cells, the T-cell receptor must interact with the foreign antigen and with a transplantation antigen. Thus the T cell recognizes the viral antigen in the context of a particular transplantation antigen, a process called H-2 restriction. Cells expressing the same viral antigen but a different transplantation antigen

Michael Steinmetz is at the Basel Institute for Immunology, Basel, Switzerland, and Leroy Hood is professor of biology and chairman of the Division of Biology, California Institute of Technology, Pasadena 91125.

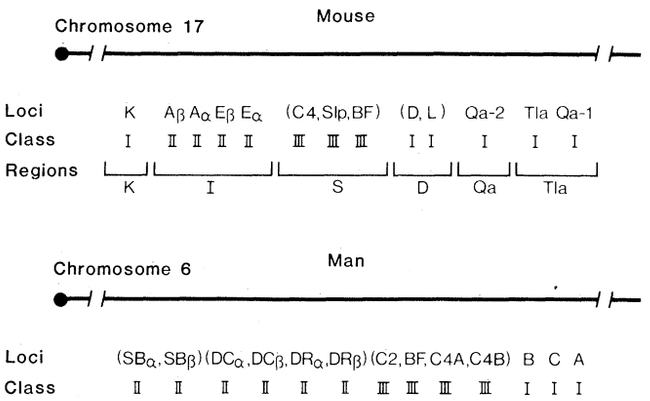

Fig. 1. Genetic maps of the mouse and human major histocompatibility complexes. The order of loci given in brackets is not known. The centromere is located to the left. For other explanations see text.

will not be killed by the same cytotoxic T cells. The functions of the Qa and TL antigens are unknown.

The class II molecules also function as restricting elements, but for regulatory T cells, which have the capacity to help or suppress a cellular or humoral immune response against foreign antigens (*9*). As we shall see subsequently, the isolation of class I and class II genes has given us the tools to begin to dissect the interactions between the restricting element, the foreign antigen, and the T-cell receptor.

Class I Genes

Structure. Class I complementary DNA (cDNA) clones were isolated by hybrid selection, primer extension, and cross-hybridization with cDNA's from a different species (*6*). The class I cDNA's then were used to screen genomic libraries constructed in lambda and cosmid vectors. The typical structure of a mouse class I gene is given in Fig. 3. All mouse class I genes that have been completely sequenced [K^b (*10*), K^d (*11*), K^k (*12*), L^d (*13*, *14*), and the Qa gene 27.1 (*15*)] are split into eight exons. The exons correlate with the structural domains of the transplantation antigen. The first exon encodes the leader or signal peptide; exons 2, 3, and 4 encode the three external domains; exon 5 encodes the transmembrane region; and exons 6, 7, and 8 encode the cytoplasmic domain. At least three different patterns of RNA splicing occur at the 3′ end of mouse class I genes (*6*, *11*). First, for the L^d gene the last intron is 139 base pairs (bp) long and is located between positions 1 and 2 of the last sense codon. Second, for the K^d and K^b genes the acceptor splice site in this intron is located 27 nucleotides upstream from the site used for L^d. Third, a cDNA clone, pH-2II, has been isolated which derives from a class I messenger RNA (mRNA) that has retained this last intron entirely.

The class I proteins encoded by these differentially spliced mRNA's will therefore differ at their carboxyl terminals. It is not known whether a single class I gene can also be spliced in two or more alternative ways. In this regard, it is interesting that the two human class I genes analyzed to date have the same exon-intron organization as the mouse genes but contain only two cytoplasmic exons (*16*).

Organization. Compared to the limited number of class I molecules identified serologically, it was surprising to find that the mouse contains about 40 class I genes. Thirty-six distinct class I genes have been isolated from a cosmid library constructed from DNA of the inbred BALB/c mouse strain (*17*). These genes have been ordered into 12 class I gene clusters which range in size from 40 to 215 kbp of DNA and contain between one and eight class I genes (*18*). A similar screen of a C57Bl/10 cosmid library has yielded seven clusters ranging in size from 40 to 120 kb with a total of about 20 class I genes (*19*). From a comparison of genomes examined by the Southern blotting technique it is unlikely that there is a twofold difference in class I genes between BALB/c and C57Bl/10 mice.

Congenic mice, a series of inbred strains with different MHC's superimposed on the same genetic background, and recombinant congenic mice, obtained from intra-MHC recombinant strains, have permitted the class I gene clusters to be assigned to one of the four MHC regions, K, D, Qa, and Tla, which encode class I molecules (Fig. 4). By using single- or low-copy probes from each of the gene clusters, restriction enzyme site polymorphisms could be identified in various inbred strains and correlated with corresponding serologic polymorphisms for the four regions in recombinant congenic mice (*19*, *20*). Four significant observations emerged from this analysis. First, all 36 class I genes isolated from BALB/c DNA map to the MHC of the mouse (*20*). This is surprising because most other multigenic families have pseudogenes that have been translocated to other chromosomal regions. Second, only 5 of the 36 class I genes from BALB/c and only 3 from the 20 class I genes from C57Bl/10 map to the H-2 complex of the MHC (*19*, *20*). Indeed, it is likely that only three of the five BALB/c genes and only two of the three C57Bl/10 genes are functional as restricting elements. In contrast, 31 of the 36 BALB/c genes and 17 of the 20 C57Bl/10 genes map to the Qa and Tla regions and hence the vast majority of the class I genes are located distal to the classical H-2 complex.

Third, a great deal of polymorphism in the flanking regions has been noted in the three BALB/c gene clusters contained in the H-2 complex, whereas very little polymorphism has been noted in these sequences in nine BALB/c gene clusters contained in the Qa and Tla regions (*20*). This restriction site polymorphism in the flanking sequences of class I genes correlates with the corresponding sequence diversity and the polymorphism of the genes contained in these respective gene clusters. It also explains why only limited differences are seen when different haplotypes are compared by Southern blot analysis with class I probes. Since most of the class I genes detected map to the Qa and Tla regions which show limited sequence variability, it appears that the restriction enzyme site polymorphism is primarily generated by just the five class I genes mapping to the H-2 complex. Thus, the MHC contains areas that show high sequence divergence and other areas that are more conserved. This will be an important point in the discussion of the I region genes. Fourth, the single- and low-copy probes from the BALB/c gene clusters have been used to count homologous fragments in the DNA's from various inbred strains of mice (*20*). These types of analyses suggest that the number of fragments detected can vary by approximately 25 percent, thus suggesting that in various inbred strains of mice the corresponding class I genes may expand or contract in number, presumably by unequal crossing-over. Most of the expansion or contraction appears to occur in the Qa and Tla regions.

Although much less information is available about the organization of class I genes in humans, it is generally concluded from Southern blot analyses as well as from cloning studies that humans have 20 to 40 class I genes. Of these, only two genes have been completely sequenced, both of which appear to be pseudogenes (*16*). Well-characterized human recombinants together with their

parental haplotypes are not avaiable, but x-ray–induced deletion mutants which have lost various portions of the HLA complex on chromosome 6 have been used for the mapping of the human genes. In fact, a recent analysis of such a deletion mutant with class I DNA probes has shown that some human class I genes map distal to the A locus to a region corresponding to the mouse Qa and Tla regions (*21*).

Polymorphic restriction fragments analyzed by Southern blots and correlated with certain HLA genes can be used for HLA typing (*22*). In addition, polymorphic restriction sites will allow us to study the association between HLA type and susceptibility to certain diseases.

Expression. DNA-mediated gene transfer of class I genes into cells of a different genetic background has permitted the identification of most of the serologically defined gene products using specific monoclonal antibodies. Thus, the K^b (*23*), K^{bm1} (*24*), K^d (*25*), K^k (*12*), D^b (*19*), D^d (*25*, *26*), L^d (*25*, *26*), Qa-2 (*25*), and two Tla genes (*25*) have been identified. Moreover, the observation that the concentrations of endogenously expressed transplantation antigens in mouse L cells remain approximately the same before and after class I gene transfer permitted the use of an assay for cell-surface β_2-microglobulin to determine whether class I gene products for which there are no available serologic reagents were being expressed (*25*). It appears that at least 15 unidentified class I genes in the BALB/c mouse can be expressed and thus may increase levels of β_2-microglobulin upon transfer into mouse L cells. All 15 of these novel genes are found in the Qa and Tla regions (Fig. 4). Thus, there may be a large number of class I antigens that are perhaps, like the Qa and TL antigens, preferentially expressed on distinct subsets of lymphocytes and indeed in the future these gene products may be useful in identifying functional subsets of lymphocytes. Several laboratories are now in the process of preparing specific antisera to the novel gene products in order to determine their patterns of expression. Other class I genes that do not increase levels of β_2-microglobulin on the cell surface might encode secreted (*27*), cytoplasmic, or pseudo class I molecules.

Human class I genes have also been transfected into mouse L cells and successfully expressed in association with mouse β_2-microglobulin. These analyses have permitted the A2, A3, B7, B40, and CW3 class I genes to be identified (*28*).

Generation of diversity. Protein and DNA sequence analyses of alleles of the transplantation antigens have demonstrated that most of the sequence variation between alleles occurs in the first and second external domains (*6*). The variation found within these two polymorphic domains between alleles is up to 20 percent at the amino acid sequence level. Within these domains there appear to be at least three regions (positions 62 to 83, 92 to 121, and 135 to 157) that are "hypervariable" (Fig. 5A). The functional implications of these allelic differences are unknown.

The origin of this diversity is unknown. An ideal system for analyzing mechanisms for generating diversity is that of the mutant transplantation antigens at the K locus of the b haplotype mouse. A series of 12 K^b mutants has been analyzed biochemically with several surprising results (*29*). First, seven of the mutants exhibit multiple amino acid substitutions rather than the single residue substitutions so characteristic of other mutant proteins such as hemoglobin and cytochrome *c*. Second, these multiple substitutions are generally tightly clustered in a single location. Third, several independently derived but identical mutants have been observed. Recently, the K^{bm1} mutant gene has been analyzed at the DNA level (*24*). The striking observation is that the mutant gene is identical to its wild-type counterpart, apart from seven nucleotide substitutions which occur within a span of 13 bases (Fig. 5B). Furthermore, it has been shown that there is a distinct class I gene in the Qa region of the b haplotype that contains the mutant sequence at the same place where it is identical in sequence to K^{bm1} over a stretch of 51 nucleotides (*30*). Presumably this second gene interacted with the K^b gene through a gene conversion event to generate the bm1 mutant gene. Thus, it is attractive to speculate that a fundamental mechanism for rapidly generating the extensive variability characteristic of class I alleles is gene conversion from one class I gene to a second in the gene family. The many class I genes located in the Qa and Tla

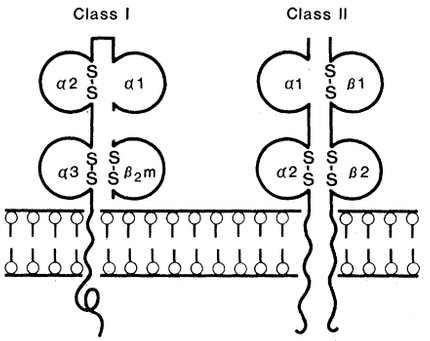

Fig. 2. Structures of class I and class II molecules. These models emphasize the structural homologies between class I and class II molecules. *S-S* indicates disulfide bridges in external domains of class I and class II molecules and β_2-microglobulin ($\beta_2 m$).

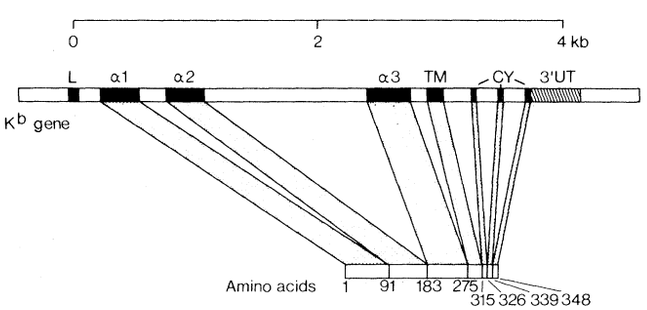

Fig. 3. H-2K^b gene structure (*10*). Exons are shown as closed boxes and labeled *L* (leader or signal peptide), *α1*, *α2*, *α3* (the three external domains), *TM* (transmembrane domain), *CY* (cytoplasmic domain), *3'UT* (3′ untranslated region). Amino acid numbers refer to the location of introns in the K^b gene.

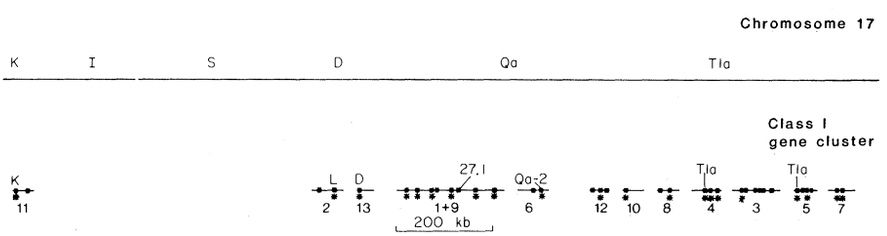

Fig. 4. The location of 36 class I genes in the MHC of the BALB/c mouse (*17*, *20*). Regions of the MHC are shown with the class I gene clusters (labeled 1 to 13) that have been mapped to them. Clusters 1 and 9 have recently been shown to overlap (*18*). Clusters 5 and 12 map either to the Qa or Tla region. Class I genes (squares) encoding serologically defined molecules are labeled accordingly; those that encode molecules which associate with β_2-microglobulins at the cell surface are indicated by asterisks. 27.1 is a class I gene that has been sequenced and mapped into the Qa region (*15*).

regions therefore might constitute a reservoir of donor sequences that can be used to alter the class I genes of the K and D loci. Such unidirectional gene conversion events affecting short DNA sequences (smaller than 50 bp) could indeed generate a large number of different class I sequences (*31*). It will be interesting to determine whether gene conversion is a fundamental mechanism for the diversification of other multigene families as well.

Function. Virus-specific cytotoxic T cells can be generated that recognize cells that have been infected with different viruses. These killer T cells can then be used to analyze class I restricting elements in mouse L cells transformed with various cloned class I genes and infected with the appropriate virus. With killer T cells prepared against lymphocytic choriomeningitis virus (LCMV), vesicular stomatitis virus (VSV), or influenza virus and mouse L cells transformed individually with cloned class I genes, several interesting observations have been made. First, only the L^d gene product works as a restricting element in LCMV infections (*32*). The same is true for the VSV infection (*33*). In contrast, with influenza virus infection it appears that in BALB/c mice both the K^d and L^d (*34*) and in C57B1/10 mice both the K^b and D^b molecules (*35*) function as restricting elements. Second, exon shuffling experiments, in which hybrid genes were constructed between the L^d and D^d genes, have led to the conclusion that only the first or second (or both) external domains participate in the restricted recognition of antigen by the T-cell receptor. This is true for virus-specific cytotoxic T cells, directed against LCMV (*36*), VSV (*36, 37*), and influenza virus (*34*), and allospecific cytotoxic T lymphocytes (*36, 37*). In contrast, some allospecific antibodies will also recognize the third external domain (*36, 38*). Third, in LCMV or VSV infections the L^d molecule may have its cytoplasmic domain deleted or altered and still be capable of functioning as a restriction element in T cell–mediated cytotoxic killing (*39*). It appears therefore that the cytoplasmic domain does not serve as an important binding site for viral cell-surface antigens (at least in the system described above) and that it is not required for the lysis of the cell. The function of the cytoplasmic domain of the class I molecules remains uncertain.

Recombination. Truncated or partial class I genes can be introduced into mouse L cells by DNA-mediated gene transfer generating hybrid genes with the frequency of one to a few percent of that of a transfection by intact class I genes, presumably by double recombination or gene conversion (*40*). A series of truncated L^d genes has been constructed. From virtually any starting combination of exons, class I antigens could be expressed containing the L^d serological determinants provided that the exon encoding the second external domain was present. The observation that truncated genes may be reconstituted in the class I gene family stands in striking contrast to the efforts to reconstitute partial genes in a variety of other examples. The question arises as to whether the ability to reconstitute truncated genes is unique to the class I gene family, is a property of most complex multigene families, or is in fact a feature of all eukaryotic genes and could be detected with appropriately sensitive assay systems. If indeed the last explanation is correct, insertion of particular genes into specific regions of the eukaryotic chromosome may be possible. Efforts are now under way to characterize the hybrid class I genes and their products to see what types of mechanisms are compatible with their generation. Recently, the reconstitution of truncated genes was confirmed with human class I genes (*41*).

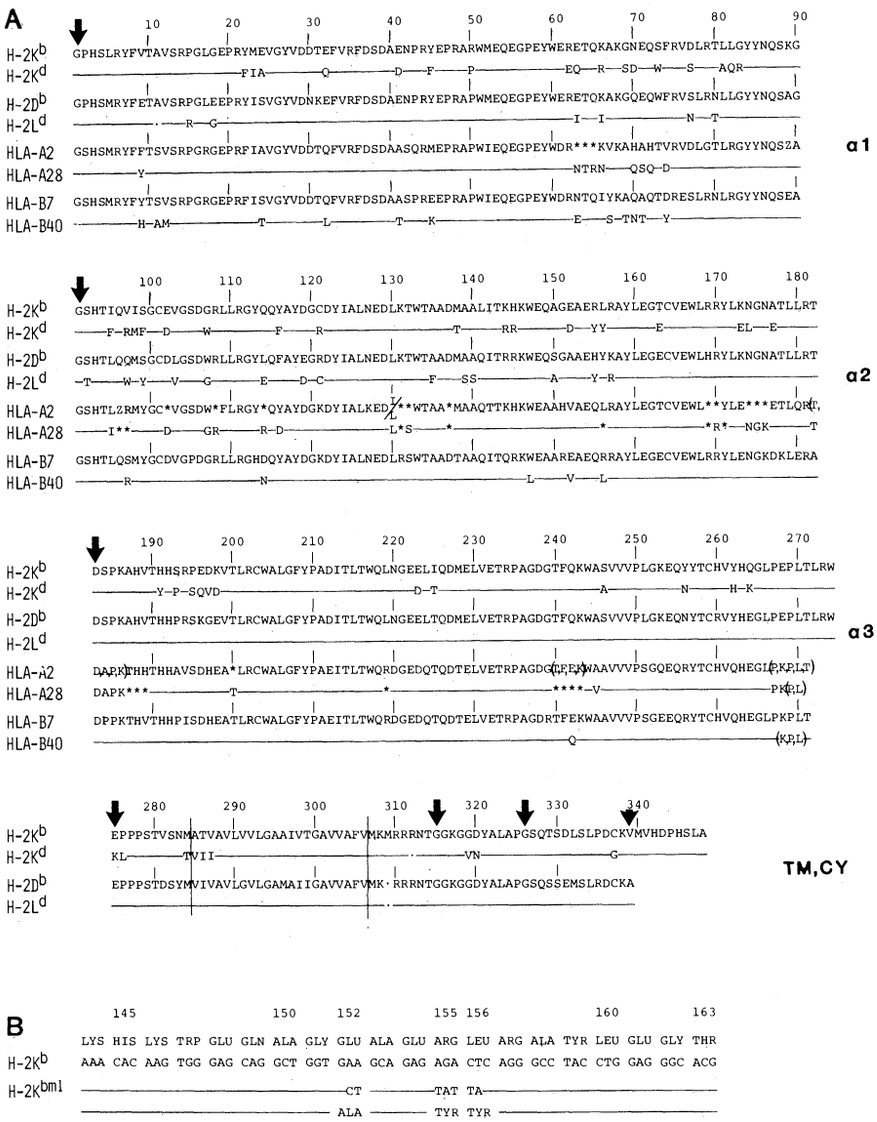

Fig. 5. (A) Sequence diversity between mouse and human class I alleles. Amino acid sequences are compared between the H-2K^b (*10*) and H-2K^d (*11*) antigens, between the H-2D^b (*63*) and H-2L^d (*13*) antigens which are presumably encoded by alleles (*6*), between the HLA-A2 and HLA-A28 antigens (*64*), and between the HLA-B7 and HLA-B40 antigens (*65*). Solid lines indicate identity to the first sequence in a group of alleles; the dot indicates a deleted residue; asterisks correspond to nonassigned positions; and arrows mark the locations of introns in the K^b gene. The portion of the molecules that appears to span the membrane is indicated by vertical lines. The single-letter code has been used for amino acids: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine, M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine; Z, glutamic acid or glutamine. (B) Comparison of the K^b and K^{bm1} nucleotide sequences (*24*). The K^{bm1} sequence is identical to the K^b sequence except for seven nucleotides in the second external domain.

Class II Genes

Structure. Techniques similar to those described for the class I genes have been used to isolate class II cDNA's (*6*). These cDNA's have been used to screen genomic libraries for the isolation of class II genes.

The DNA sequences are known for the mouse A_α^d (*42*), A_α^k (*43*), A_β^d (*44, 45*), A_β^k (*44*), A_β^b (*44, 46*), E_α^d (*47*), E_α^k (*48*), and E_β^d (*49*) genes as well as for the human DC_α (*50*), DC_β (*51*), DR_α (*52*), and DR_β (*53*) genes. Sequence comparisons demonstrate that the human DR genes are analogous to the mouse I-E genes and the human DC genes are the counterparts to the mouse I-A genes. So far the mouse homologs to the human SB genes have not been identified, although the SB_β gene appears to cross-hybridize best with the mouse $E_{\beta 2}$ gene described in the next section (*54*).

The mouse E_α and the human DR_α genes are split into five exons, whereas the mouse A_β and E_β genes are split into six exons (Fig. 6). As for the class I genes, there is a correlation between exon-intron organization and domain structure of the class II genes and proteins. The structural difference between the α and β genes occurs at the 3′ end: α genes have only one intervening sequence in the 3′ untranslated region whereas the A_β and E_β genes show two intervening sequences located in the cytoplasmic region but no intervening sequence in the 3′ untranslated region. It is interesting that the DC_β gene—the human homolog to the mouse A_β gene—has lost exon 5 encoding part of the cytoplasmic domain probably because of a nonfunctional splice acceptor site (*51*). Mouse and human class I and class II genes show similar structural variations at their 3′ ends. It is not clear whether these changes are important for different effector functions of the molecules and for distinct interactions with cytoskeletal proteins or whether they simply reflect evolutionary variations of a nonfunctional carboxyl terminus.

Organization. The organization of the class II genes in the I region of the BALB/c mouse has been studied by chromosomal "walking" procedures (*55*). In brief, a DR_α cDNA clone was used to isolate four overlapping cosmid clones containing the homologous mouse E_α gene. Restriction fragments free of repetitive DNA sequences were then isolated from the ends of the cloned region and were used to screen the cosmid library a second time to obtain additional clones at either end of this region. This process was then carried out repeatedly and led to the isolation of about 270 kb of DNA from the I region of the BALB/c mouse (Fig. 7). All of the serologically defined class II genes of the mouse are contained within this region—E_α and E_β as well as A_α and A_β. In addition, two more class II genes were found, $A_{\beta 2}$ and $E_{\beta 2}$. The $A_{\beta 2}$ gene appears to be a single isolated exon (*46*) and accordingly might be a pseudogene, although it has not been ruled out that the missing coding sequences are located further upstream and downstream but are only distantly related to the class II sequences cloned so far. Northern blot hybridization experiments with the DNA sequences around the $A_{\beta 2}$ gene should clarify this point. The $E_{\beta 2}$ gene appears to be a similar situation.

The α and β genes have opposite orientations to one another. This suggests that the β-β-α group was an evolutionary subunit which duplicated once in the case of the mouse to give the I-A and I-E subregion genes and perhaps twice in the case of humans to give the DR, DC, and SB sets of genes. It should be pointed out that the organization of the class II genes in a second mouse strain, the AKR mouse, from which this region has been cloned (*55*), is the same as that of the BALB/c mouse.

The organization of the human class II α and β genes is not known to the same extent as in the mouse. It is currently believed that there are five or six genes (one DR_α, three or four DC_α-related, and one SB_α) and seven β genes (three DR_β, two DC_β, and two SB_β) (*53, 56*).

The I-J paradox. The I-J subregion has been mapped by serological analyses of intra-I region recombinant strains to a position between the I-A and I-E subregions (Fig. 7). The I-J subregion is of functional interest because it appears to encode I-J polypeptides which are subunits of T cell suppressor factors. However, a molecular analysis of this region suggests that the I-J polypeptides probably are not encoded here.

Comparison of the molecular map of the I region of the mouse with the genetic map has confirmed the location of A_α and A_β genes in the I-A subregion and the location of the E_α gene in the I-E subregion (Fig. 7). The E_β gene, however, is located with its 5′ end in the I-A subregion and with its 3′ end in the I-E subregion. A region of 2 kb which spans part of the long intervening sequence between exons β1 and β2 and part of the β2 exon has so far not been assigned to either the I-A or I-E subregion (*57*). In

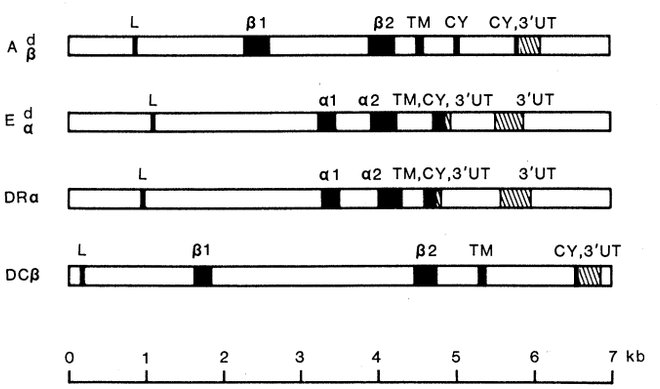

Fig. 6. Exon-intron organization of the mouse A_β^d and E_α^d and the human DR_α and DC_β genes. The mouse E_β gene has the same exon-intron organization as the A_β gene. *L*, signal or leader peptide; *α1*, *α2*, *β1*, *β2*, external domains; *TM*, transmembrane domain; *CY*, cytoplasmic domain; *3′UT*, 3′ untranslated region.

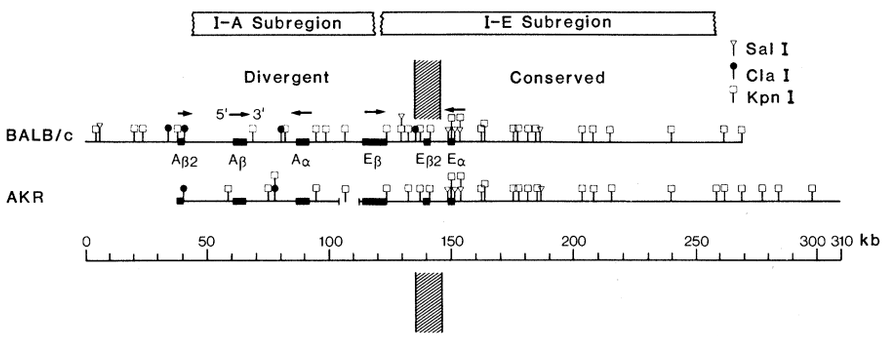

Fig. 7. Organization of the class II genes in the I region of BALB/c and AKR mice (*55*). The vertical bar shows the approximate location of a breakpoint between two regions of high and low restriction enzyme site variability. The gap in the AKR map indicates a region that is not represented in the isolated cosmid clones. The cloned portion of the I region has so far not been linked to the K region to the left and the S region to the right. The recombination separating the I-A from the I-E subregion has occurred in the middle of the E_β gene (see text). Note that in our previous map (*55*) a portion of the restriction map was inadvertently exchanged between the BALB/c and AKR mouse.

theory, this region could encode the antigenic determinants of the I-J polypeptide. Serological analyses of a suppressor factor secreted by suppressor T cells specific for lactate dehydrogenase B has revealed the presence of E_β and I-J determinants on the same polypeptide chain suggesting that the I-J polypeptide is encoded by a differentially spliced E_β mRNA (*58*). Northern blot experiments, however, carried out under highly sensitive conditions, have failed to identify RNA transcripts from the E_β gene region in a variety of I-J positive suppressor T cells (*59*). DNA sequence analysis of the E_β gene region in intra-I region recombinant mouse strains which encode serologically distinguishable I-J polypeptide chains will help to resolve the I-J paradox. Such analyses are now under way.

Expression. Mouse and human class II genes have been introduced into mouse L cells and successfully expressed at the cytoplasmic and cell-surface levels (*60*). A third polypeptide, denoted the invariant (Ii) chain, has been found associated with class II polypeptides in the cytoplasm and has been postulated to be involved in the transport of class II polypeptides from the cytoplasm to the cell surface (*61*). It is unlikely that this hypothesis is correct because mouse L cells express very little of the invariant chain compared to B cells where class II genes are normally expressed. L cells transformed with human class II genes have been isolated which express as much human Ia antigens on the cell surface as B cells. Furthermore, cotransfer of mouse class II genes and the gene for the invariant chain into mouse L cells does not increase the relatively low cell-surface concentration of Ia antigens on the surface of some transformed cells although the Ii gene is transcribed at high levels. Recently, cell-surface expression levels in mouse L cells comparable to those seen in mouse B cells have been obtained with the transformed A_α and A_β genes (*60*). These transformed L cells also are capable of presenting antigen to T-helper cell hybridomas (*60*). The successful expression of class II genes after gene transfer will now permit dissection of the function of Ia antigens by mutagenesis in vitro.

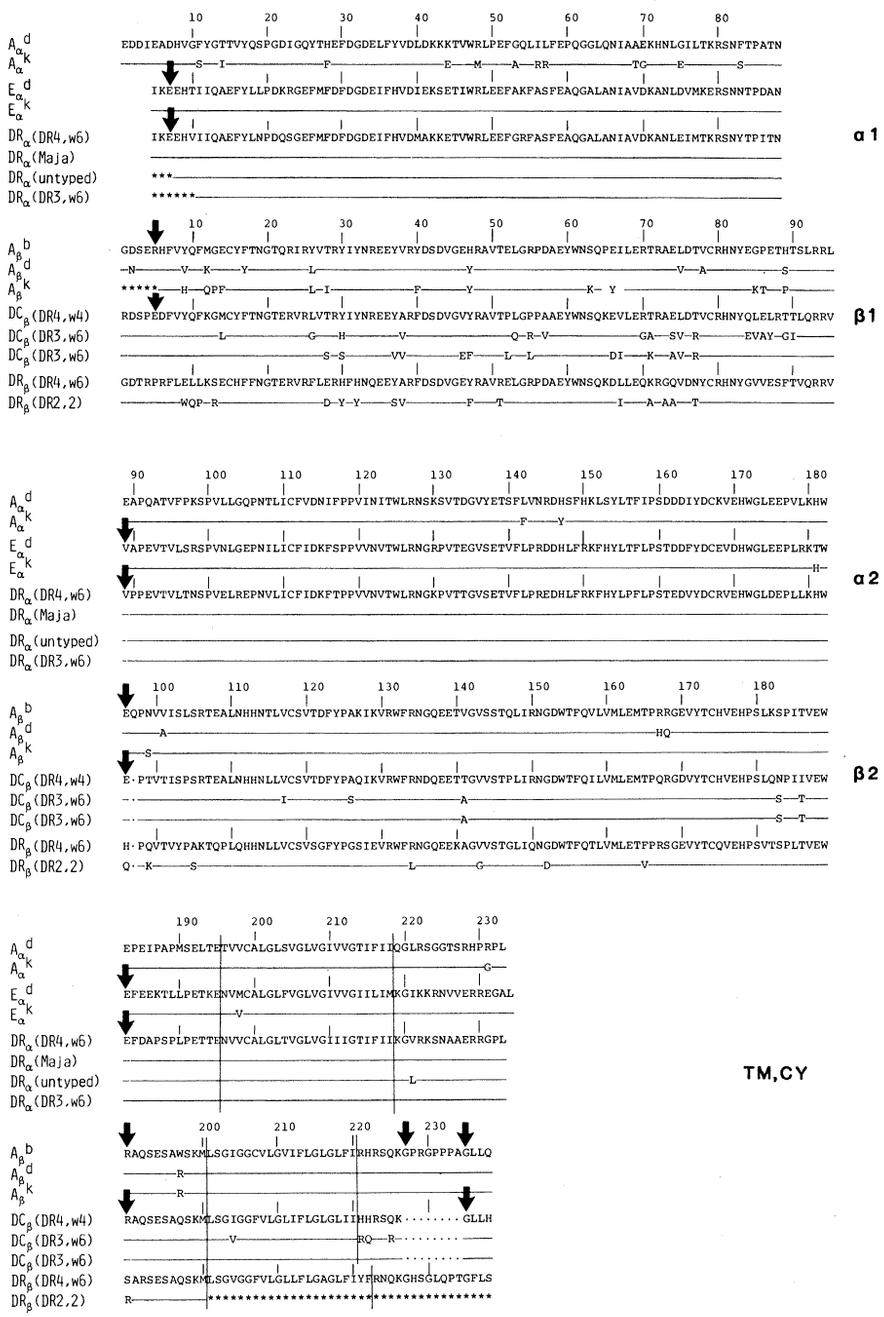

Fig. 8. Sequence diversity between mouse and human class II alleles. Amino acid sequences are compared between the A^d_α (*42*) and A^k_α (*43*) alleles; between the E^d_α (*47*) and E^k_α (*48*) alleles; between four DR_α alleles from a DR4,w6 B cell line, Maja cells, an untyped individual, and a DR3,w6 B cell line (*52*); between the A^b_β (*44, 46*), A^d_β (*44, 45*), and A^k_β (*44*) alleles; between three DC_β alleles, one from a DR4,w4 homozygous individual and two from a DR3,w6 B cell line (*51*); and between two DR_β alleles from a DR4,w6 (*53*) and a DR2,2 (*66*) B cell line. Location of introns (arrows) are indicated where known. For explanations of other symbols see legend to Fig. 5.

Polymorphism. Serological, biochemical, and more recently DNA sequence analyses have shown that the mouse A_α, A_β, and E_β genes and the human DC_α, DC_β, and DR_β genes are polymorphic, whereas the mouse E_α and the human homolog, DR_α, are conserved (*7*). Sequence comparisons reveal that the alleles of polymorphic class II loci—like those for class I alleles—are very different from each other (Fig. 8). These differences are mainly found in the α1 and β1 domains (10 to 23 percent sequence divergence at the protein level) whereas the α2, β2, transmembrane, and cytoplasmic domains are conserved. Especially interesting is the observation that the transmembrane domain between the mouse A_α and E_α chains (*43*), the human DC_α and DR_α chains (*50*), and the DC_β and DR_β chains (*53*) are more conserved than the Ig-like α2 and β2 domains (Fig. 8). The opposite appears to be true for class I genes (Fig. 5A) (*6*). Perhaps the transmembrane region of class II molecules is conserved because it exerts important molecular interactions, either between the α and β chains or with other transmembrane proteins (*7, 50*).

Comparison of the restriction maps of the I region from two mouse inbred strains (BALB/c and AKR) has revealed

that the allelic variability of the A_α, A_β, and E_β loci correlates with restriction site differences in this region whereas the restriction map around the conserved E_α locus shows very little variation between the two strains (Fig. 7). Southern blot analyses of other inbred strains confirm that the I region can be split into areas of high and low restriction site variability (*55*). Taken together with the variability of restriction sites around class I genes (*20*), it appears that the MHC can be divided into chromosomal domains that are highly divergent (K region, I-A subregion, D region) and chromosomal domains that are relatively conserved (I-E subregion, Qa, and Tla regions) (*62*). Increased frequencies of mutation in certain chromosomal areas therefore appear to have contributed to the exceptional diversity of the class I, K, and D alleles and the alleles of the class II A_β, A_α, and E_β loci. As discussed above, a mechanism for diversifying coding regions in class I genes appears to be gene conversion. Whether gene conversion events are also important for diversifying class II genes is not clear.

An analysis of the nine intra-I region recombinants has demonstrated that all nine recombination events have occurred within a distance of 8 kb or less at the E_β gene locus (*55*). This hot spot of recombination correlates approximately with the boundary between the conserved and divergent domains (Fig. 7). One wonders whether there is a connection between the propensity for recombination in this region and the transition from a conserved to a variable chromosomal domain. Perhaps certain aspects of chromosomal structure can promote or suppress the genetic mechanisms responsible for generating sequence diversity and focus recombination to a confined region on the chromosome.

Evolution of MHC Genes

The external domains closest to the cell membrane of the class I and class II molecules are homologous to one another, to β_2-microglobulin, and to the constant region domains of immunoglobulins (*6*). These homologies suggest that the genes encoding the class I, class II, β_2-microglobulin, and antibody gene families were derived from a common ancestor. One wonders whether additional gene families encoding cell-surface antigens will be discovered that also show homology to the genes of the MHC and the immunoglobulins. It is possible that the genes encoding the T-cell receptor molecules are also members of this supergene family.

References and Notes

1. J. Klein, *Biology of the Mouse Histocompatibility-2 Complex* (Springer-Verlag, New York, 1975); *Science* **203**, 516 (1979); G. D. Snell, J. Dausset, S. Nathenson, *Histocompatibility* (Academic Press, New York, 1976).
2. L. Flaherty, in *The Role of the Major Histocompatibility Complex in Immunology*, M. E. Dorf, Ed. (Garland, New York, 1981), p. 33; M. J. Soloski, J. W. Uhr, L. Flaherty, E. S. Vitetta, *J. Exp. Med.* **153**, 1080 (1981); J. Michaelson *et al.*, *Transplant. Proc.*, in press.
3. K. Yokoyama, E. Stockert, L. J. Old, S. Nathenson, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 7078 (1981).
4. T. Cotner, H. Mashimo, P. C. Kung, G. Goldstein, J. L. Strominger, *ibid.*, p. 3858; C. Terhorst *et al.*, *Cell* **23**, 771 (1981).
5. D. A. Shackelford, J. F. Kaufman, A. J. Korman, J. L. Strominger, *Immunol. Rev.* **66**, 133 (1982); R. Tosi, N. Tanigaki, D. Centis, G. B. Ferrara, D. Pressman, *J. Exp. Med.* **148**, 1592 (1978); C. K. Hurley, S. Shaw, L. Nadler, S. Schlossman, J. D. Capra, *ibid.* **156**, 1557 (1982).
6. L. Hood, M. Steinmetz, B. Malissen, *Annu. Rev. Immunol.* **1**, 529 (1982).
7. J. F. Kaufman, C. Auffray, A. J. Korman, D. A. Shackelford, J. L. Strominger, *Cell*, in press.
8. R. M. Zinkernagel and P. C. Doherty, *Adv. Immunol.* **27**, 51 (1980).
9. Z. Nagy, C. N. Baxevanis, N. Ishii, J. Klein, *Immunol. Rev.* **60**, 59 (1981).
10. E. Weiss *et al.*, *EMBO J.* **2**, 453 (1983).
11. S. Kvist, L. Roberts, B. Dobberstein, *ibid.*, p. 245.
12. S. Kvist, personal communication.
13. K. W. Moore, B. T. Sher, Y. H. Sun, K. A. Eakle, L. Hood, *Science* **215**, 679 (1982).
14. G. A. Evans, D. H. Margulies, R. D. Camerini-Otero, K. Ozato, J. G. Seidman, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 1994 (1982).
15. M. Steinmetz *et al.*, *Cell* **25**, 683 (1981).
16. M. Malissen, B. Malissen, B. R. Jordan, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 893 (1982); P. A. Biro, J. Pan, A. K. Sood, R. Kole, V. B. Reddy, S. M. Weissman, *Cold Spring Harbor Symp. Quant. Biol.* **47**, 1082 (1982).
17. M. Steinmetz, A. Winoto, K. Minard, L. Hood, *Cell* **28**, 489 (1982).
18. In the original paper (*17*) 13 class I gene clusters were identified. Recently it was shown by J. Rogers and K. R. Willison [*Nature (London)* **304**, 549 (1983)] that clusters 1 and 9 overlap.
19. R. A. Flavell *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **47**, 1067 (1982).
20. A. Winoto, M. Steinmetz, L. Hood, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3425 (1983).
21. H. T. Orr and R. DeMars, *Nature (London)* **302**, 534 (1983).
22. H. M. Cann, L. Ascanio, P. Paul, A. Marcadet, J. D. Dausset, D. Cohen, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1665 (1983).
23. A. L. Mellor *et al.*, *Nature (London)* **298**, 529 (1982).
24. E. H. Weiss *et al.*, *ibid.* **301**, 671 (1983); D. H. Schulze *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2007 (1983).
25. R. S. Goodenow, M. McMillan, M. Nicolson, B. T. Sher, K. Eakle, N. Davidson, L. Hood, *Nature (London)* **300**, 231 (1982).
26. D. H. Margulies *et al.*, *J. Immunol.* **130**, 463 (1983).
27. M. Kress, D. Cosman, G. Khoury, G. Jay, *Cell* **34**, 189 (1983).
28. J. A. Barbosa, M. E. Kamarck, P. A. Biro, S. M. Weissman, F. H. Ruddle, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 6327 (1982); F. A. Lemonnier *et al.*, *Immunogenetics* **18**, 65 (1983).
29. R. Nairn, K. Yamaga, S. G. Nathenson, *Annu. Rev. Genet.* **14**, 241 (1980); L. R. Pease, D. H. Schulze, G. M. Pfaffenbach, S. G. Nathenson, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 242 (1983).
30. R. A. Flavell, personal communication.
31. F. Brégère, *Biochimie* **65**, 229 (1983).
32. A. Örn *et al.*, *Nature (London)* **297**, 415 (1982).
33. J. Forman, R. S. Goodenow, L. Hood, R. Ciavarra, *J. Exp. Med.* **157**, 1261 (1983).
34. C. Reiss, G. A. Evans, D. H. Margulies, J. G. Seidman, S. J. Burakoff, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2709 (1983).
35. A. R. M. Townsend, P. M. Taylor, A. L. Mellor, B. A. Askonas, *Immunogenetics* **17**, 283 (1983).
36. I. Stroynowski *et al.*, in preparation.
37. K. Ozato, G. A. Evans, B. Shykind, D. H. Margulies, J. G. Seidman, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2040 (1983).
38. G. A. Evans, D. H. Margulies, B. Shykind, J. G. Seidman, K. Ozato, *Nature (London)* **300**, 755 (1982).
39. M. C. Zuniga, B. Malissen, M. McMillan, P. R. Brayton, S. S. Clark, J. Forman, L. Hood, *Cell*, in press.
40. R. S. Goodenow, I. Stroynowski, M. McMillan, M. Nicolson, K. Eakle, B. T. Sher, N. Davidson, L. Hood, *Nature (London)* **301**, 388 (1983).
41. S. M. Weissman and P. Lingell, personal communication.
42. M. M. Davis, D. I. Cohen, E. A. Nielsen, M. Steinmetz, W. E. Paul, L. Hood, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
43. C. O. Benoist, D. J. Mathis, M. R. Kanter, V. E. Williams II, H. O. McDevitt, *ibid.* **80**, 534 (1983).
44. E. Choi, K. McIntyre, R. N. Germain, J. G. Seidman, *Science* **221**, 283 (1983).
45. M. Malissen, T. Hunkapiller, L. Hood, *ibid.*, p. 750.
46. D. Larhammar *et al.*, *Cell*, in press.
47. J. McNicholas, M. Steinmetz, T. Hunkapiller, P. Jones, L. Hood, *Science* **218**, 1229 (1982); J. J. Hyldig-Nielsen *et al.*, *Nucleic Acids Res.* **11**, 5055 (1983).
48. D. J. Mathis, C. O. Benoist, V. E. Williams II, M. R. Kanter, H. McDevitt, *Cell* **32**, 745 (1983).
49. H. Saito, R. A. Maki, L. K. Clayton, S. Tonegawa, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
50. C. Auffray, A. Korman, M. Roux-Dosseto, R. Bono, J. L. Strominger, *ibid.* **79**, 6337 (1982).
51. D. Larhammar *et al.*, *ibid.*, in press; D. Larhammar *et al.*, *Hum. Immunol.*, in press; D. Larhammar *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 3687 (1982).
52. D. Larhammar *et al.*, *Cell* **30**, 153 (1982); J. S. Lee *et al.*, *Nature (London)* **299**, 750 (1983); A. J. Korman, C. Auffray, A. Schamboeck, J. L. Strominger, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 6013 (1982); H. K. Das, S. K. Lawrence, S. M. Weissman, *ibid.* **80**, 3543 (1983).
53. E. O. Long, C. T. Wake, J. Gorski, B. Mach, *EMBO J.* **2**, 389 (1983).
54. B. Mach, personal communication.
55. M. Steinmetz *et al.*, *EMBO J.* **300**, 35 (1982); M. Steinmetz *et al.*, unpublished results.
56. C. Auffray, J. Kuo, R. DeMars, J. L. Strominger, *Nature (London)* **304**, 174 (1983); E. O. Long and B. Mach, personal communication.
57. J. A. Kobori, A. Winoto, E. Gibb, L. Hood, J. McNicholas, *Mol. Cell. Immunol.*, in press.
58. Z. Ikezawa *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
59. M. Kronenberg *et al.*, *ibid.*, in press.
60. B. Malissen, M. Steinmetz, M. McMillan, M. Pierres, L. Hood, *Nature (London)*, in press; C. Rabourdin-Combe and B. Mach, *ibid.* **303**, 670 (1983); B. Malissen *et al.*, in preparation.
61. S. Kvist, K. Wiman, L. Claesson, P. A. Peterson, B. Dobberstein, *Cell* **29**, 61 (1982).
62. For a discussion of the I-B and I-C subregions, see J. Klein, F. Figueroa, Z. A. Nagy, *Annu. Rev. Immunol.* **1**, 119 (1983).
63. A. A. Reyes, M. Schöld, R. B. Wallace, *Immunogenetics* **16**, 1 (1982); W. L. Maloy and J. E. Coligan, *ibid.*, p. 11.
64. J. A. Lòpez de Castro, J. L. Strominger, D. M. Strong, H. T. Orr, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 3813 (1983).
65. J. A. Lòpez de Castro, R. Bragado, D. M. Strong, J. L. Strominger, *Biochemistry* **22**, 3961 (1983).
66. H. Kratzin *et al.*, *Hoppe-Seyler's Z. Physiol. Chem.* **362**, 1665 (1981).

67. We thank R. Flavell, S. Kvist, P. Lingell, E. Long, B. Mach, and S. Weissman for communication of unpublished results; J. Kaufman, D. Larhammar, J. Lòpez de Castro, R. Maki, J. Michaelson, and Z. Nagy for preprints; and G. Farmer, J. Kaufman, J. Kobori, and S. Nathenson for comments on the manuscript. This work was supported in part by grants from the National Institutes of Health. The Basel Institute for Immunology was founded and is supported by F. Hoffmann–La Roche Limited Company, Basel, Switzerland.