ORGANIZATION AND EVOLUTION OF THE CLASS I GENES IN THE MURINE MAJOR HISTOCOMPATIBILITY COMPLEX

Thesis by

Yi Henry Sun

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1986

(submitted January 23, 1986)

To Fei, my wonderful wife.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Leroy Hood, for providing an intellectually stimulating environment and the never-ending resources. I am grateful to James Ou, Charlie Rice, and Kevin Moore for patiently teaching me the basic molecular biology techniques in my first year at Caltech. I am also grateful to my two principal collaborators: Michael Steinmetz, who suggested to me the dm1 project, and to Steve Hunt, who was a great help during the preparation of this thesis. My special thanks to Ray Owen, whose office in 07 Kerckhoff is always warm, welcoming, and quite often provides delicious cake. Thanks to the many people in Lee's lab, particularly Ellen Kraig, Roger Perlmutter, Joan Klotz, Carol Readhead, Rick Barth, Jerry Siu, Astar Winoto, Mitch Kronenberg, Ulf Landegren, and Karyl Minard, who made the years at Caltech so enjoyable and interesting.

I am also grateful to "the girls in the kitchen": Bertha Jones, Ria DeBruyn, and Jessie Walker for keeping the lab running smoothly. I would like to thank my PC, Connie Katz, and Cathy Elkins for the preparation of this thesis.

I would like to thank the Earle C. Anthony Fellowship for financial support and the Weigle Fund for support in the preparation of this thesis.

Lastly, my greatest thanks to my parents, who made all this possible.

ABSTRACT

This thesis contains studies of the organization and evolution of the class I gene family in the murine major histocompatibility complex (the H-2 complex).

The first chapter presents the molecular characterization of the H-2^{dm1} mutation. The mutant gene is shown to be formed by the fusion of the 5' part of the D^d gene and the 3' part of the L^d gene, with the region in between deleted.

Chapter Two describes the results of chromosome walking experiments and presents a molecular map of 500 kb of cloned DNA, which links the *H-2D* and *Qa* regions and contains five *D* region and eight *Qa* region class I genes.

Chapter Three presents the DNA sequences of the transmembrane exon from 20 class I genes, and the use of 23 low copy-number flanking-region probes to detect homology between the regions containing each gene. The sequence comparison and the hybridization patterns indicate that multiple recombinational events, notably gene duplication and gene conversion, have occurred during the evolution of this large gene family.

Chapter Four presents a rapid method of restriction site mapping of cosmids and plasmids. The method was developed due to the need of mapping a large number of clones during the course of this study.

TABLE OF CONTENTS

Page

Acknowledgements iii

Abstract iv

Introduction 1

Chapter One Molecular Basis of the dm1 Mutation in the Major Histocompatibility Complex of the Mouse: A *D/L* Hybrid Gene 27

Chapter Two Organization and Evolution of D Region Class I Genes in the Mouse Major Histocompatibility Complex 43

Chapter Three Class I Genes of the Murine Major Histocompatibility Complex Are A Patchwork Created by Multiple Recombinational and Conversional Events 84

Chapter Four A Rapid Method for Restriction Site Mapping of Cosmids and Plasmids 123

Conclusion 139

INTRODUCTION

1. History of H-2

The major histocompatibility complex (MHC) was first discovered in the mouse by two seemingly unrelated traits controlled by the locus: the presence of an antigen on erythrocytes (Gorer, 1936) and the susceptibility to tumor transplants (Gorer, 1937). It was later shown that the cell surface antigen was responsible for the susceptibility to tumor and tissue transplants. Because the antigen was responsible for tissue compatibility, it was designated as one of the histocompatibility antigens, and the gene coding for the antigen was serially numbered *H-2* (histocompatibility-2) (Gorer, Lyman and Snell, 1948). As more traits and the responsible genes were mapped into this locus, it became clear that it contained more than one gene and was called the *H-2* complex. Many histocompatibility genes have been identified, but the immune response elicited by these genes is much weaker and has a much slower onset course than the response elicited by the *H-2* complex. Hence the *H-2* is called the *major* histocompatibility complex. For a detailed historical account on the early years of H-2 research, see Klein (1975).

2. Genetic Organization

The *H-2* complex has been mapped genetically to chromosome 17. Recently, *in situ* hybridization to pachytene chromosomes definitively placed the location of *H-2* to the junction of bands B and C of chromosome 17 (Lader et al., 1985).

The *H-2* complex can be subdivided into four major regions, *K*, *I*, *S* and *D*, and the *I* region can be further divided into subregions *I-A* and *I-E* (Figure 1). The *K* and *D* regions contain genes encoding the classical transplantation antigens, namely K, D and L. The *I* region contains genes encoding the immune response-associated or Ia antigens, namely Aα, Aβ, Eα and Eβ. The *S* region contains genes encoding certain components (C2, Bf, Slp, and C4) of the complement system and a steroid biosynthesis enzyme, 21-hydroxylase (Chaplin et al., 1983; Amor et al., 1985; White et al., 1984; Steinmetz, personal communication).

Although a vast number of phenotypic traits (more than 60; see Klein, 1978) have been reported to be controlled by the *H-2* complex, most, if not all, of them can be accounted for by the protein products encoded by the genes described above (Klein et al., 1981; 1983). Evidence for the existence of additional regions or genes has been previously reported. But in view of recent studies, most of these can be discarded. The *G, I-B,* and *I-C* regions can be accounted for without the need to postulate new regions (see Klein et al., 1983). The existence of *I-J* is not supported by molecular cloning of the region (Steinmetz et al., 1982; Kronenberg et al., 1983; Kobori et al., 1984). The gene (*I-N*) encoding the T cell-specific Iat.W41 antigen has been mapped to between the *K* and *I-A* region (Hayes and Bach, 1980), but based on the molecular mapping of the recombination points of the recombinant mouse strains used in the original mapping, the gene does not map to this region (Steinmetz et al., 1986). Genes coding for a group of low molecular weight proteins have also been mapped into the region between *K* and *I-A* (Monaco and McDevitt, 1982). The gene (*Neu-1*) encoding neuraminidase-1 was mapped to between the *I* and *D* region (Figueroa et al., 1982). These latter two loci have not been shown to be erroneously mapped, and may have been accidentally entrapped in the MHC.

Adjacent to the *H-2D* region is the *Tla* (Thymus leukemia antigen) complex, which includes the *Qa-2, Tla* and the *Qa-1* regions (Figure 1), and encodes the Qa and TL lymphoid differentiation antigens. These antigens are structurally similar to the H-2 transplantation antigens, and as discussed in later sections, their genes are also closely related to the *H-2K* and *D* region genes. Therefore, the usage of the term "MHC" has been expanded to include both the *H-2* and *Tla* complexes. Additional loci, *Qa-3, Qa-4,* and *Qa-5* have been proposed (Flaherty et al., 1981; Hammerling et al., 1979), but are not firmly established as separate loci, since no recombination between them has been observed.

Because the *H-2* is a gene complex, the particular assembly of alleles of all of the *H-2* genes on the same chromosome is conveniently termed its *H-2* haplotype, and is denoted by a superscript of lower case letters and numbers. For example, the *H-2* of the BALB/c inbred mouse is designated the "d" haplotype, written as $H\text{-}2^d$, and contains the K^d gene which encodes the K^d antigen. The BALB/c *Qa-2*, *Tla*, and *Qa-1* loci have the a, c, and b alleles, respectively.

The MHC has been found in all mammals and vertebrates studied to date. In each case adequately analyzed, a close linkage of class I and class II genes is found. And in the few cases that have been looked at, class II genes are also associated with the MHC. Comparable molecular studies have been done with the human MHC (*HLA*), and are consistent with the H-2 studies. Clones of class I gene have been isolated from pig (Singer et al., 1982; Satz et al., 1985) and rat (Diamond et al., 1986).

3. The Class I antigens

The H-2 transplantation antigens, namely, K, D, and L, are codominantly expressed, and are present on the surface of virtually all somatic cells. They are integral membrane glycoproteins with an apparent molecular weight of 45 kilodaltons (kd) and are associated noncovalently with a 12 kd polypeptide, β_2-microglobulin (β_2m), which is encoded in chromosome 2. The H-2 transplantation antigens are extremely polymorphic, with about 100 alleles per locus (Klein and Figueroa, 1981).

The transplantation antigens are composed of three external regions (α1, α2 and α3, also called N, C1 and C2), each of which is about 90 amino acid residues in length, a transmembrane region of about 40 residues, and a cytoplasmic region of about 30 residues (Figure 2). Glycosylation sites have been identified on all three external regions. The α2 and α3 regions each have a centrally located disulfide bridge spanning about 60 residues (for review, see Ploegh et al., 1981; Hood et al.,

1983). Protein sequence comparison showed that the first two external regions ($\alpha 1$, $\alpha 2$) are relatively variable and the third region ($\alpha 3$) is highly conserved. Proteolytic digestion studies have shown that the alloantigenic determinants are primarily located on the first two external regions ($\alpha 1$ and $\alpha 2$), while the third external region ($\alpha 3$) is responsible for the binding to β_2m (Yokoyama and Nathenson, 1983). Exon-shuffling experiments have also shown that the allo-antigenic determinants recognized by antibodies and cytotoxic T cells are primarily located on the $\alpha 1$ and $\alpha 2$ regions (see Stroynowski et al., 1985), consistent with their high level of variability. The divergence of the $\alpha 3$ region is limited probably due to the functional constraint of interacting with the β_2-microglobulin. The transmembrane region contains a hydrophobic stretch of about 25 residues, which presumably traverses the membrane, and is followed by several positively charged residues, which have been suggested to serve as an anchor on the cytoplasmic side of the membrane (Ploegh et al., 1981). Recent studies (Zuniga et al., 1983; Zuniga and Hood, 1986) have shown that, although their presence enhances the cell surface expression, these charged amino acids are not required for the protein to be membrane-bound, and that they may play a role in the intracellular transport of the protein. The cytoplasmic region has been suggested as being responsible for the transduction of the lysis signal when the class I protein, in association with foreign antigens on the target cell, is recognized by the cytotoxic T cells (Berke and Clark, 1982). However, studies have shown that cells expressing a class I molecule devoid of the cytoplasmic tail can still be effectively lysed by the cytotoxic T cells (Zuniga et al., 1983).

An X-ray crystallographic study of the human tranplantation antigens suggested that the $\alpha 1$ and $\alpha 2$ regions jointly form a structural domain, and the $\alpha 3$ region forms a domain with β_2m (Bjorkman et al., 1985). This domain structure is similar to the domain structure of immunoglobulins and also similar in the size of

the domain and the position of the disulfide bridge. Sequence comparison also shows a statistically highly significant level of homology among the class I α3 region, β_2m, and the immunoglobulin constant region. Similar domain structure and sequence homology were also found in the class II molecules, the poly-Ig receptor, the T-cell accessory molecules T4 and T8, Thy-1, and a neuronal antigen MRC-OX2, and have been interpreted as evidence for the common evolutionary origin of these proteins (see Kronenberg et al., 1986).

The Qa-2,3, TL and Qa-1 antigens are similar to the H-2 transplantation antigens in several aspects. First, they are also integral membrane glycoproteins, with molecular weights of 40-45 kd, and also are associated with β_2-microglobulin. Second, an antiserum against Qa-1 crossreacts with H-2K (Figueroa et al., 1983). Third, cytotoxic T cells generated against Qa-2 are not H-2 restricted (Forman et al., 1982; Klein et al., 1983), a feature specific to H-2 class I or class II antigens. As described in a later section, their genes are also homologous to the H-2 transplantation genes and probably share the same evolutionary origin. Therefore, these antigens, along with the H-2 transplantation antigens, are called the class I antigens, and their genes belong to the class I gene family. They are, however, different from the H-2 class I antigens in two important aspects. First, they have a limited tissue distribution. TL is expressed only on thymocytes, activated T cells, and certain leukemias (Flaherty, 1981) and Qa-2 is expressed on T and B lymphocytes, hematopoietic stem cells and granulocytic-macrophage precursors (Kincade et al., 1980). Second, they have a limited polymorphism (Tewarson et al., 1983).

4. The Class I Genes

The MHC contains at least two multigene families. The class I gene family includes genes located in the *H-2K*, *H-2D*, and the *Qa-Tla* region, which encode the H-2 transplantation antigens, the Qa-2,3-like antigens, and the thymus

leukemia (TL) antigen. The class II gene family contains at least seven homologous genes located in the *H-2I* region. A third class of genes encodes the complement components (*C2*, *Bf*, *Slp*, *C4*) and they are located in the *H-2S* region. Two genes encoding the 21-hydroxylase have also been shown to reside in the *H-2S* region. It is the class I gene family that is the subject of this thesis.

Based on homology to a H-2 class I cDNA probe (Steinmetz et al., 1981), more than 30 class I genes have been isolated from the BALB/c mouse (Steinmetz et al., 1982). These were grouped into 13 clusters by overlapping restriction maps. By examining the DNA of various congenic and recombinant congenic mouse strains, these clusters were mapped with respect to the genetic regions of the MHC (Winoto et al., 1983). A large number of the genes were unexpectedly mapped to the *Qa* and *Tla* regions. By transfecting mouse L cells with the various class I genes and assaying with a panel of antisera and monoclonal antibodies, genes coding for the K, D, L, Qa-2,3, and TL antigens have been identified (Goodenow et al., 1982). DNA sequences of the genes encoding the K (Kvist et al., 1983), D (Sher et al., 1985), L (Moore et al., 1982; Evans et al., 1982) and TL (Fisher et al., 1985) antigens have been determined. Similar molecular cloning and characterization have also been done with mouse of the $H\text{-}2^b$ and $H\text{-}2^k$ haplotypes (Weiss et al., 1984; Steinmetz et al., 1984).

The class I genes are in general split into eight exons, corresponding to the structural components of the protein (Figure 3). The first exon encodes the leader or signal peptide, exons 2, 3 and 4 encode the three external regions, exon 5 encodes the transmembrane region, and exons 6-8 together encode the cytoplasmic region. Alternative RNA splicing in the first exon has been shown to occur in the K^d gene (Lalanne et al., 1983), and the same alternative splice sites can also be identified in the D^d and L^d genes. Multiple lines of evidence also suggested the existence of alternative splicing at the 3' end (see Hood et al.,

1983). No gross DNA rearrangement of the class I genes can be detected in cells expressing them as is seen for the immunoglobulin genes (Steinmetz et al., 1982).

5. H-2 Mutants

Mouse H-2 mutations that cause the class I antigens to gain and/or lose antigenic determinant(s) can be selected by graft rejection. About 30 mutants have been identified by screening approximately 100,000 mice (Klein, 1978; Mckenzie et al., 1977a; see also Sun et al., 1985). Studies of these mutants have yielded valuable information. The dm2 mutant mouse, which lacks the L^d antigen, was used to define the *L* locus (Mckenzie et al., 1977b; Hansen et al., 1977). The mutants were important in the study of the function of the H-2 class I antigens (McKenzie et al., 1977a; Klein, 1978). Interestingly, a large fraction (22/30) of the mutations occurred in the *K* gene. Fourteen K^b mutants have been analyzed biochemically and all show limited structural differences from the parental K antigens (Nairn et al., 1980). In each case, the amino acid changes and the corresponding nucleotide changes are clustered (see Hood et al., 1983). Gene conversion has been suggested to be the mechanism for generating these mutations (Pease et al., 1983; Weiss et al., 1983). In fact, a potential donor gene has been identified that contains the exact DNA sequence of the converted short stretch in the $H\text{-}2K^{bm1}$ mutant gene (Mellor et al., 1983). Six mutations map into the *H-2D* region and show more extensive alterations than the K mutants (see Sun et al., 1985).

6. Polymorphism and Function

One of the most striking features of the H-2 class I antigens is their extremely high genetic polymorphism. This is unusual not only in the large number of alleles, but also in the relatively low frequency of any individual allele. In general, other polymorphic loci have a few alleles and one of the alleles is considerably more frequent than the others (see Dausset, 1981; Klein, 1979). Because of the high polymorphism and the low frequency of each allele, more than

90 percent of the wild mice in a population are heterozygous at the *H-2* class I loci. This polymorphism has been suggested (Bodmer, 1973) not to be a true genetic polymorphism of a single locus but results from the existence of many loci, each encoding for one of the "allelic" forms. The polymorphism derives from regulation of which of the genes is to be expressed. From the molecular cloning studies, there are clearly not enough genes to account for the hundred or so allelic forms.

Although the H-2 class I molecules were originally identified for their role in transplantation, hence the name transplantation antigens, their physiological role is probably participation in the recognition of foreign antigens by cytotoxic T lymphocytes (T_C). Cytotoxic T cells recognize foreign antigens only when the antigen is presented on the target cell surface together with the proper class I molecule, a phenomenon termed "H-2 restriction" (Zinkernagel and Doherty, 1979; Schwartz, 1985). The class I molecule is called the restriction element for T_C. Similarly, the class II molecules, originally associated with the immune responses to certain simple antigens, have now been shown to act as restriction elements for the helper T cells.

It is generally agreed that the extremely high polymorphism of the H-2 class I molecule is related and important to their function as restriction elements, but it is uncertain as to how it is related. The final answer may not come until the precise role of the H-2 antigens in the recognition of antigens by the T cell receptor is understood. The most convincing theory at the present time can be described as the following. Since the foreign antigens have to be recognized in association with the H-2 class I or class II molecules, polymorphism ensures that for every antigen there will always be some alleles of the H-2 antigens that can interact with it and be recognized by the immune system. Polymorphism therefore has a selective advantage for the population, and since most of the population is

heterozygous for the class I loci, it also increases the ability of each individual to defend against infections. Indeed, some H-2 class I alleles of the laboratory mouse strains do not allow the recognition of certain viral antigens (Bubbers et al., 1978).

There is no evidence to indicate that the Qa and TL antigens can act as restriction elements (Klein et al., 1983). The function, if any, of these antigens is unknown. Although they have been called differentiation antigens, it is only because they have restricted tissue distribution and can be used as markers for different differentiation stages or cell lineages. No role in differentiation has been found.

7. Problems and Approaches

Several problems were of interest to me. I will describe each problem and the approaches I have taken to address these problems.

(1) Are the class I genes regulated in their expression, and if so, how?

Six class I antigens, H-2K, H-2D, H-2L, Qa-1, Qa-2 and TL, have been definitively defined by serological, genetic, and biochemical methods. But the number of class I genes isolated far exceeds the number of these defined antigens. How many of these genes are expressed and where and when are they expressed? How is their expression regulated?

One mutation in the H-2 class I transplantation antigens caught my attention and suggested a possibility to identify and characterize a regulatory mechanism. The dm1 mutation, unlike all the other H-2 mutations, affects both the D^d and L^d genes. Immunoprecipitation studies first showed that not only are both the D^d and L^d antigens altered, but an additional molecule of 39 kd was present in the mutant (Wilson et al., 1981). This complex phenotype suggested to me the possibility of a regulatory mutation where the expression of a set of class I genes was turned off, and another set of class I genes was turned on.

Chapter One presents the molecular characterization of the dm1 mutation. The mutation, to my disappointment, is the result of the fusion of the *H-2D* and *L* genes, forming a functional hybrid gene. No regulatory mechanism is apparently involved. It was later reported (Hunt et al., 1985) that only one H-2D- or H-2L-related molecule can be identified in the dm1 mutant; the other two molecules previously reported were experimental artifacts.

(2) The structure-function relationship of the class I antigens.

Using hybrid class I genes constructed by exon-shuffling methods, it has been possible to determine which regions of the H-2 class I antigen are important for the interaction with antibodies and cytotoxic T lymphocytes (see Stroynowski et al., 1985). The dm1 mutation provides a natural hybrid gene. Whereas the hybrid genes artificially constructed to date have exchanged complete exons, the dm1 *D/L* hybrid gene has a hybrid α2 exon, allowing a finer functional dissection within a single external region. The results and discussions of these studies are presented also in Chapter One.

(3) How many class I genes are in the $H\text{-}2D^d$ region?

In addition to the well-characterized D^d and L^d protein, three class I antigens, R^d, M^d, and $L2^d$, have been detected serologically and mapped genetically to the $H\text{-}2D^d$ region (see Hansen et al., 1983), while only three class I genes (D^d, L^d, and a gene located 5' to L^d) have been cloned from the $H\text{-}2^d$ mouse. The molecular basis of these additional class I antigens is not clear. They may be encoded by genes other than the D and L genes, some of which may not have been isolated, or may be derived from the D or L gene via an alternative RNA splicing pathway or with different posttranslational modification.

To determine how many class I genes are located in the D^d region, chromosomal walking technique was used to clone the entire D^d region. The molecular organization of class I genes in the D^d region and its comparison with those of the

D^b and D^k regions are presented in Chapter Two.

Studies of the molecular basis of the dm1 and dm2 mutations also provided information regarding the origin of the additional class I antigens. These are discussed in Chapters One and Three.

(4) What are the evolutionary processes that generated the class I multigene family? What are the mechanisms responsible for generating the high degree of polymorphism in the *H-2* class I genes?

To study the evolutionary process that generated the large number of genes in the class I gene family, it is desirable to compare sequences of the same region from as many genes as possible. We have used a simple method to sequence the transmembrane exon and its 5' and 3' introns from most of the BALB/c class I genes. From sequence comparisons and data obtained by using many sequences, both from the coding region of the various class I genes and from regions flanking the genes, as hybridization probes, the relatedness of the class I genes was studied. The results are presented in Chapter Three.

(5) During the course of study, due to the need of restriction mapping many cosmid clones, a method for rapid restriction site mapping was developed. This method is presented in Chapter Four.

REFERENCES

Amor, M., Tosi, M., Duponchel, C., Steinmetz, M., and Meo, T. (1985). Liver mRNA probes disclose two cytochrome P-450 genes duplicated in tandem with the complement *C4* loci of the mouse *H-2 S* region. Proc. Natl. Acad. Sci. USA. **82,** 4453-4457.

Berke, G., and Clark, W. R. (1982). T lymphocyte-mediated cytolysis—a comprehensive theory. 1. The mechanisim of CTL-mediated cytolysis. Adv. Exp. Med. Biol. **146,** 57-68.

Bjorkman, P. J., Strominger, J. L., and Wiley, D. C. (1985). Crystallization and X-ray diffraction studies on the histocompatibility antigens HLA-A2 and HLA-A28 from human cell membranes. J. Mol. Biol. **186,** 205-210.

Bodmer, W. F. (1973). A new genetic model for allelism at histocompatibility and other complex loci: Polymorphism for control of gene expression. Transpl. Proc. **5,** 1471-1475.

Bubbers, J. E., Chen, S., and Lilly, F. (1978). Nonrandom inclusion of H-2K and H-2D antigens in Friend virus particles from mice of various strains. J. Exp. Med. **147,** 340-351.

Chaplin, D. D., Woods, D. E., Whitehead, A. S., Goldberger, G., Colten, H. R., and Seidman, J. G. (1983). Molecular map of the murine *S* region. Proc. Natl. Acad. Sci USA. **80,** 6947-6951.

Dausset, J. (1981). The major histocompatiblity complex in man. Science **213,** 1469-1474.

Diamond, A. G., Windle, J. M., Butcher, G. W., Winoto, A., Hood, L., and Howard, J. C. (1986). Identification and expression of genes encoding rat class I and II MHC molecules from two genomic cosmid libraries. Transplant. Proc., in press.

Evans, G. A., Margulies, D. H., Camerini-Otero, R. D., Ozato, K., and Seidman, J. G. (1982). Structure and expression of a mouse major histocompatibility antigen gene, *H-2L*^{d}. Proc. Natl. Acad. Sci. USA **79**, 1994-1998.

Figueroa, F., Klein, D., Tewarson, S., and Klein, J. (1982). Evidence of placing the *Neu-1* locus within the mouse *H-2* complex. J. Immunol. **129**, 2089-2093.

Figueroa, F., Zaleska-Rutcznyska, Z., Kusnierczyk, P., and Klein, J. (1983). Crossreactivity between Qa-1 region and H-2K antigens. Transplantation **35**, 391-393.

Fisher, D. A., Hunt S.W. III, and Hood, L. (1985). Structure of a gene encoding a murine thymus leukemia antigen, and organization of *Tla* genes in the BALB/c mouse. J. Exp. Med. **162**, 528-545.

Flaherty, L. (1981). *Tla*-region antigens. In *The Role of the Major Histocompatibility Complex in Immunobiology*, ed. M. Dorf (Garland, New York), pp. 33-57.

Forman, J., Trial, J., Tonkonogy, S., and Flaherty, L. (1982). The QA2 subregion controls the expression of two antigens recognized by H-2-unrestricted cytotoxic T cells. J. Exp. Med. **155**, 749-767.

Goodenow, R. S., McMillan, M., Nicolson, M., Sher, B. T., Eakle, K., Davidson, N., and Hood, L. (1982). Identification of the class I genes of the mouse major histocompatibility complex by DNA-mediated gene transfer. Nature **300**, 231-237.

Gorer, P. A. (1936). The detection of antigenic differences in mouse erythrocytes by the employment of immune sera. Brit. J. Exp. Pathol. **17**, 42-50.

Gorer, P. A. (1937). The genetic and antigenic basis of tumor transplantation. J. Pathol. Bacteriol. **44**, 691-697.

Gorer, P. A., Lyman, S., and Snell, G. D. (1948). Studies on the genetic and antigenic basis of tumor transplantation: linkage between a histocompatibility gene and "fused" in mice. Proc. Roy. Soc. (London) **B135**, 499-505.

Hammerling, G. J., Hammerling, U., and Flaherty, L. (1979). Qa-4 and Qa-5, new murine T-cell antigens governed by the *Tla* region and identified by monoclonal antibodies. J. Exp. Med. **150**, 108-116.

Hansen, T. H., Cullen, S. E., Melvold, R., Kohn, H. I., Flaherty, L., and Sach, D. H. (1977). Mutation in a new *H-2* associated histocompatibility gene closely linked to *H-2D*. J. Exp. Med. **145**, 1550-1558.

Hansen, T. H., Ozato, K., and Sachs, D. H. (1983). Heterogeneity of *H-2D* region associated genes and gene products. Adv. Immunol. **34**, 39-70.

Hayes, C. E. and Bach, F. H. (1980). *I-N*, a newly described *H-2 I* subregion between *K* and *I-A*. J. Exp. Med. **151**, 481-485.

Hood, L., Steinmetz, M., and Malissen, B. (1983). Genes of the major histocompatibility complex of the mouse. Ann. Rev. Immunol. **1**, 529-568.

Hunt, P., Burnside, S. S., and Sears, D. W. (1985). Evidence of only one H-2D/L-related glycoprotein in H-2^{dm1} mutant cells. J. Immunogenetics, **12**, 33-44.

Kincade, P., Flaherty, L., Lee, G., Watanabe, T., and Michaelson, J. (1980). Presence of the Qa-2 alloantigen on functional lymphoid myeloid and stem cells in adult mice. J. Immunol. **124**, 2879-2882.

Klein, J. (1975). ***Biology of the Mouse Histocompatibility-2 Complex.*** Springer-Verlag, New York.

Klein, J. (1978). *H-2* mutations: their genetics and effect on immune functions. Adv. Immunol. **26**, 55-146.

Klein, J. (1979). The major histocompatibility complex of the mouse. Science **203**, 516-521.

Klein, J., and Figueroa, F. (1981). Polymorphism of the mouse *H-2* loci. Immunol. Rev. **60**, 23-57.

Klein, J., Juretic, A., Baxevanis, C. N., and Nagy, Z. A. (1981). The traditional and a new version of the mouse *H-2* complex. Nature **291**, 455-460.

Klein, J., Figueroa, F., and Nagy, Z. A. (1983). Genetics of the major histocompatibility complex: The final act. Ann. Rev. Immunol. **1**, 119-142.

Kobori, J. A., Winoto, A., McNicholas, J., and Hood, L. (1984). Molecular characterization of the recombination region of six murine major histocompatibility complex (MHC) *I*-region recombinants. J. Mol. Cell Immunol. **1**, 125-131.

Kronenberg, M., Steinmetz, M., Kobori, J., Kraig, E., Kapp, J., Pierce, C. W., Sorensen, C. M., Suzuki, G., Tada, T., and Hood, L. (1983). RNA transcripts for I-J polypeptides are apparently not encoded between the *I-A* and *I-E* subregions of the murine major histocompatibility complex. Proc. Natl. Acad. Sci. USA. **80**, 5704-5708.

Kronenberg, M., Siu, G., Hood, L. E., and Shastri, N. (1986). The molecular genetics of the T-cell antigen receptor. Ann. Rev. Immunol., in press.

Kvist, S., Roberts, L., and Dobberstein, B. (1983). Mouse histocompatibility genes: structure and organisation of a K^d gene. EMBO J. **2**, 245-254.

Lader, E., Clark, B. T., Jhanwar, S. C., Chaganti, R. S. K., and Bennett, D. (1985). Definitive chromosomal location of the *H-2* complex by *in situ* hybridization to pachytene chromosomes. Immunogenetics **22**, 49-54.

Lalanne, J.-L., Transy, C., Guerin, S., Darche, S., Meulien, P. and Kourilsky, P. (1985). Expression of class I genes in the major histocompatibility complex: identification of eight distinct mRNAs in DBA/2 mouse liver. Cell **41**, 469-478.

Lalanne, J.-L., Cochet, M., Kummer, A.-M., Gachelin, G., and Kourilsky, P. (1983). Different exon-intron organization at the 5' part of a mouse class I gene is used to generate a novel H-2K^{d}-related mRNA. Proc. Natl. Acad. Sci. USA **80**, 7561-7565.

McKenzie, I. F. C., Pang, T., and Blanden, R. V. (1977a). The use of H-2 mutants as models for the study of T cell activation. Immunol. Rev. **35**, 181-230.

McKenzie, I. F. C., Morgan, G. M., Melvold, R. W., and Kohn, H. I. (1977b). BALB/c-H-2^{db}: a new H-2 mutant in BALB/cKh that identified a locus associated with the D region. Immunogenetics **4**, 333-347.

Mellor, A. L., Weiss, E. H., Ramachandran, K., and Flavell, R. A. (1983). A potential donor gene for the bm1 gene conversion event in the C57BL mouse. Nature **306**, 782-795.

Monaco, J. J., and McDevitt, H. O. (1982). Identification of a fourth class of proteins linked to the murine major histocompatibility complex. Proc. Natl. Acad. Sci. USA **79**, 3001-3005.

Moore, K. W., Sher, B. T., Sun, Y. H., Eakle, K. A., Hood, L. (1982). DNA sequence of a gene encoding a BALB/c mouse L^{d} transplantation antigen. Science **215**, 679-682.

Pease, L. R., Schulze, D. H., Pfaffenback, G. M., and Nathenson, S. G. (1983). Spontaneous H-2 mutants provide evidence that a copy mechanism analogous to gene conversion generates polymorphism in the major histocompatibility complex. Proc. Natl. Acad. Sci. USA 80, 242-246.

Ploegh, H. L., Orr, H. T., and Strominger, J. L. (1981). Major histocompatibility antigens: The human (HLA-A, -B, -C) and murine (H-2K, H-2D) class I molecules. Cell **24**, 287-299.

Nairn, R., Yamaga, K., and Nathenson, S. G. (1980). Biochemistry of the gene products from murine MHC mutants. Ann. Rev. Genet. **14**, 241-277.

Satz, M. L., Wang, L.-C., Singer, D.S., and Rudikoff, S.(1985). Structure and expression of two porcine genomic clones encoding class I MHC antigens. J. Immunol. **135**, 2167-2175.

Schwartz, R. H. (1985). T-Lymphocyte recognition of antigen in association with gene products of the major histocompatibility complex. Ann. Rev. Immunol. **3**, 237-261.

Sher, B. T., Nairn, R., Coligan, J. E., Hood, L. E. (1985). DNA sequence of the mouse *H-2D*^{d} transplantation antigen gene. Proc. Natl. Acad. Sci. USA **82**, 1175-1179.

Singer, D. S., Camerini-Otero, R. D., Satz, M. L., Osborne, B., Sachs, D., and Rudikoff, S. (1982). Characterization of a porcine genomic clone encoding a major histocompatibility antigen: Expression in mouse L cells. Proc. Natl. Acad. Sci. USA **79**, 1403-1407.

Steinmetz, M., Moore, K. W., Frelinger, J. G., Sher, B. T., Shen, F.-W., Boyse, E. A., and Hood, L. (1981). A pseudogene homologous to mouse transplantation antigens: transplantation antigens are encoded by eight exons that correlate with protein domains. Cell **25**, 683-692.

Steinmetz, M., Winoto, A., Minard, K., and Hood, L. (1982). Clusters of genes encoding mouse transplantation antigens. Cell **28**, 489-498.

Steinmetz, M., Malissen, M., Hood, L., Orn, A., Maki, R. A., Dastoornikoo, G. R., Stephan, D., Gibb, E., and Romaniuk, R. (1984). Tracts of high or low sequence divergence in the mouse major histocompatibility complex. EMBO J. **3**, 2995-3003.

Steinmetz, M., Stephan, D., and Lindahl, K. F. (1986). Gene organization and recombinational hotspots in the murine major histocompatiblity complex. Cell, in press.

Stroynowski, I., Clark, S., Henderson, L. A., Hood, L., McMillan, M., and Forman, J. (1985). Interaction of α1 with α2 region in class I MHC proteins contributes determinants recognized by antibodies and cytotoxic T cells. J. Immunol. **135**, 2160-2166.

Sun, Y. H., Goodenow, R. S., and Hood, L. (1985). Molecular basis of the dm1 mutation in the major histocompatibility complex of the mouse: a D/L hybrid gene. J. Exp. Med. **162**, 1588-1602.

Tewarson, S., Zaleska-Rutczynska, Z., Figueroa, F., and Klein, J. (1983). Polymorphism of *Qa* and *Tla* loci of the mouse. Tissue Antigens **22**, 204-212.

Weiss, E., Golden, L., Zakut, R., Mellor, A., Fahrner, K., Kvist, S., and Flavell, R. A. (1983). The DNA sequence of the $H\text{-}2K^b$ gene: evidence for gene conversion as a mechanism for the generation of polymorphism in histocompatibility antigens. EMBO J. **2**, 453-462.

Weiss, E. H., Golden, L., Fahrner, K., Mellor, A. L., Devlin, J. J., Bullman, H., Tiddens, H., Bud, H., and Flavell, R. A. (1984). Organization and evolution of the class I gene family in the major histocompatibility complex of the C57BL/10 mouse. Nature **310**, 650-655.

White, P. C., Chaplin, D. D., Weis, J. H., Dupont, B., New, M. I., and Seidman, J. G. (1984). Two steroid 21-hydroxylase genes are located in the murine *S* region. Nature **312**, 465-467.

Wilson, P. H., Nairn, R., Nathenson, S. G., and Sears, D. W. (1981). Unexpected complexity of the dm1 mutation revealed in the structure of three H-2D/L-related antigens. Immunogenetics **15**, 225-237.

Winoto, A., Steinmetz, M., and Hood, L. (1983). Genetic mapping in the major histocompatibility complex by restriction enzyme site polymorphisms: Most mouse class I genes map to the *Tla* complex. Proc. Natl. Acad. Sci. USA **80**, 3425-3429.

Yokoyama, K., and Nathenson, S. G. (1983). Intramolecular organization of class I H-2 MHC antigens; localization of the alloantigenic determinants and the β_2m binding site to different regions of the H-2 K^b glycoprotein. J. Immunol. **130**, 1419-1425.

Zinkernagel, R. M., and Doherty, P. C. (1979). MHC restricted cytotoxic T cells: Studies on the biological role of polymorphic major transplantation antigens determining T cell restriction specificity, fuction and responsiveness. Adv. Immunol. **27**, 51-75.

Zuniga, M. C., Malissen, B., McMillan, M., Brayton, P. R., Clark, S. S., Forman, J., and Hood, L. (1983). Expression and function of transplantation antigens with altered or deleted cytoplasmic domains. Cell **34**, 535-544.

Zuniga, M. C., and Hood, L. E. (1986). Clonal variation in cell surface display of an H-2 protein lacking a cytoplasmic tail. J. Cell Biol. **102**, 1-10.

Figure 1. Genetic map of the murine MHC.

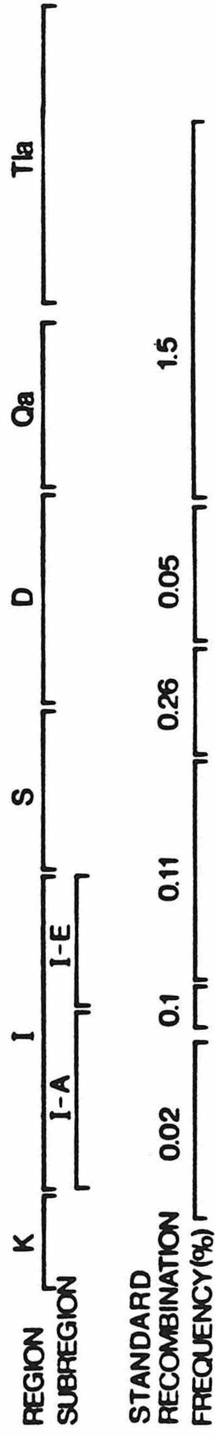

REGION
SUBREGION
K
I
S
D
Qa
Tla
I-A
I-E
STANDARD
RECOMBINATION
FREQUENCY(%)
0.02
0.1
0.11
0.26
0.05
1.5

Figure 2. Structure of a class I glycoprotein.

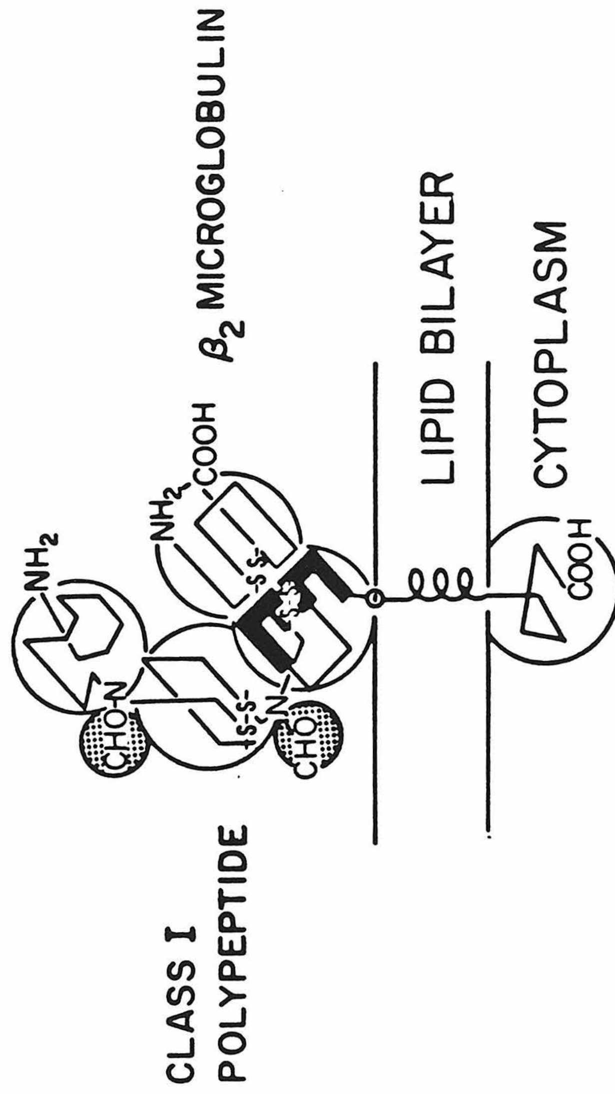
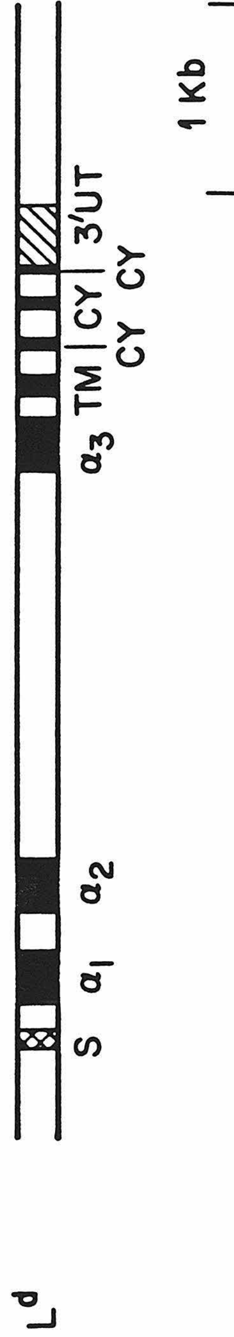

CLASS I
POLYPEPTIDE
β_2 MICROGLOBULIN
NH2
COOH
NH2
CHO
CHO
-S-S-
-S-S-
-S-S-
LIPID BILAYER
CYTOPLASM
COOH

Figure 3. Exon and intron organization of a class I gene.



L^d
S
α1
α2
α3
TM
CY
CY
CY
3'UT
1 Kb

CHAPTER ONE

MOLECULAR BASIS OF THE *dm1* MUTATION IN THE MAJOR HISTOCOMPATIBILITY COMPLEX OF THE MOUSE: A *D/L* HYBRID GENE

Published in the *Journal of Experimental Medicine*

MOLECULAR BASIS OF THE dm1 MUTATION IN THE MAJOR HISTOCOMPATIBILITY COMPLEX OF THE MOUSE: A D/L HYBRID GENE

By Y. HENRY SUN, ROBERT S. GOODENOW, AND LEROY HOOD

From the Division of Biology, California Institute of Technology, Pasadena, California 91125

The major histocompatibility complex (MHC)[1] of mammals encodes molecules that play important roles in the immune response. The H-2 region of the murine MHC contains at least three gene families. Cytotoxic T lymphocytes recognize foreign antigens in the context of the gene products from one of these families, the class I transplantation antigens (1). The transplantation antigens are integral membrane glycoproteins with a molecular weight of about 45 kilodaltons (kD). They are associated noncovalently with β_2-microglobulin and are present on the surfaces of most somatic cells.

The combination of alleles at all loci within the H-2 complex is termed its haplotype and is denoted by a superscript. The inbred BALB/c mouse has the H-2^d haplotype, and three H-2 class I transplantation antigens, K^d, D^d, and L^d, have been defined both serologically and biochemically. The genes encoding these antigens have been cloned and characterized (2). The K^d gene maps to the proximal (relative to the centromere) end of the H-2 complex, while the D^d and L^d genes map to the D region at the distal end of the H-2 complex. The presence of additional H-2^d class I antigens, e.g., M^d, R^d, has been suggested by serological analyses (3, 4). These antigens have not been well characterized.

The transplantation antigens are composed of three external regions ($\alpha 1$, $\alpha 2$, $\alpha 3$), each about 90 residues in length, a transmembrane region, and a cytoplasmic region. The genes are split into eight exons, corresponding to the structural components of the protein. The first exon encodes the leader or signal peptide, exons 2, 3, and 4 encode the three external regions, exon 5 encodes the transmembrane region, and exons 6, 7, and 8 encode the cytoplasmic region.

Mutations in the H-2 genes that cause the transplantation antigens to gain and/or lose antigenic determinant(s) can be selected by graft rejection. 30 mutant transplantation antigens have been identified (5–10). Studies of these mutants have yielded valuable information on the structure-function relationship of the class I antigens (6, 11). In most characterized mutants, the mutation can be localized to a small number of amino acid substitutions of the protein (7, 12).

This work was supported by grant AI19624 from the National Institutes of Health, Bethesda, MD. Y. Sun is supported by an Earle C. Anthony Fellowship. The present address of R. Goodenow is Department of Genetics, University of California, Berkeley, CA 94720.

[1] *Abbreviations used in this paper:* CTL, cytotoxic T lymphocyte; mAb, monoclonal antibody; MHC, major histocompatibility complex; NC, nitrocellulose; SDS, sodium dodecyl sulfate.

 J. Exp. Med. © The Rockefeller University Press · 0022-1007/85/11/1588/15 $1.00
Volume 162 November 1985 1588–1602

One mutant, dml, however, is unique in that it extensively affects both the H-2D^d and H-2L^d antigens.

The dml mutant was derived from the inbred strain B10.D2 of the H-2^d haplotype (13). Serological and transplantation studies showed the mutant is of the gain-and-loss type, i.e., it has gained and also lost antigenic determinant(s) (13–16). The mutation was mapped to the H-2D region (17, 18). The most interesting feature of this mutation is that both the H-2D^d and the H-2L^d antigens are affected (19–21). In the mutant, only a single H-2D- or -L-related class I antigen can be identified (22–25). This molecule was called the D^{dml} antigen because it shares an H-2D^d-specific serological determinant (22). The D^{dml} antigen shares only ~70% of its tryptic peptides with those of the D^d antigen, in contrast to the ≥90% peptide homology between most of the mutant K^b molecules and their wild-type counterparts, indicating a significant change that cannot be easily explained by point mutations (22). Tryptic peptide comparison and analyses with monoclonal antibodies (mAb) have suggested that the D^{dml} antigen is a hybrid protein between the D^d and L^d antigens (24).

This study investigates the basis of the dml mutation at the DNA level, and shows that the D^{dml} antigen is the product of a hybrid gene formed by fusion of the 5′ part of the D^d gene and the 3′ part of the L^d gene. The mutation probably resulted from a simple recombinational event with the region between the D^d and L^d genes deleted.

Materials and Methods

Mice. BALB/cJ and B10.D2 were obtained from The Jackson Laboratory, Bar Harbor, ME. B10.D2-H2^{dml} was kindly provided by D. Sears (University of California, Santa Barbara).

Hybridization Probes. All probes were cloned in M13 vectors. Probe 59.2A was provided by A. Winoto (25a). Probe pH-2^{d}27.51 was a gift from M. Steinmetz (Basel Institute for Immunology, Basel, Switzerland) (26). Probe 18.1B is a 0.2 kilobase (kb) Alu I fragment from a 3.5 kb Bam HI/Nru I fragment of cosmid clone 18.1 subcloned into Sma I site of M13mp8 (B. Sher, Stanford University School of Medicine, unpublished observation). Probe 59.2B is a 0.7 kb Sau 3A fragment from a 3.2 kb Sma I/Hpa I fragment of cosmid clone 59.2 subcloned into Bam HI site of M13mp8.

Since M13mp vectors contain part of the *lac* region from *E. coli*, probes labeled by nick translation or primer extension will also hybridize to the *E. coli* chromosome. For screening cosmid libraries, probes were either gel-purified inserts or were prepared as follows. Single-stranded phage DNA was labeled by primer extension using the M13 universal sequencing primer (27). After the reactions, the sample was ethanol-precipitated and resuspended in 10 mM Tris, pH 8.0, 1 mM sodium EDTA. The insert was cut out by digestion with Eco RI and Pst I at the polylinker. The sample was adjusted to 1 M ammonium acetate and passed through nitrocellulose (NC) filter that was premoistened with 1 M ammonium acetate. The insert, now fully double stranded, will pass through, while the partially single-stranded vector fragment will be bound to the NC filter. The NC filtering step could be easily done by putting a small piece of NC filter in a 0.5 ml Eppendorf tube with a hole in the bottom. This small tube was placed in a 1.5 ml Eppendorf tube. The sample was spotted onto the NC filter. The eluate was collected in the 1.5 ml tube as the double tube was spun in an Eppendorf microfuge for 30 s. Probes had a specific activity of ~10^8 cpm/μg of insert DNA.

Cosmid Library Construction and Screening. The dm1 cosmid library was constructed using Mbo I partially digested and size-selected B10.D2-H-2^{dml} liver DNA cloned into the Bgl II site of the vector pTL5 following the procedure of Steinmetz et al. (26). The library contains ~360,000 colonies distributed on 60 NC filters (137 mm). The cloning

efficiency was ~10^6 CFU (colony-forming units)/μg of size-selected insert DNA. The BALB/cJ cosmid library was similarly constructed using liver DNA from BALB/cJ mice, except that the nylon membrane, Biotrans (ICN Biomedicals, Inc., Irvine, CA), was used instead of the NC filters.

Hybridization is carried out in 5× SSPE (0.18 M NaCl, 10 mM NaH_2PO_4, pH 7.4, 1 mM EDTA), 5× Denhardt's solution, 100 μg/ml denatured, sheared salmon sperm DNA, 0.1% sodium dodecyl sulfate (SDS), 10^5 cpm/ml of each probe, at 68°C for 16–18 h. The filters were washed with 0.2× SSPE and 0.1% SDS, at 68°C. Exposures were for 8–10 h at −70°C with an intensifying screen.

Cosmid Clones. Cosmid clones 16.1, 59.2, and 49.2 were isolated from a BALB/cCum sperm DNA cosmid library and have been described (26–28). Cosmid clone 12.1 was isolated from a BALB/cJ sperm DNA cosmid library constructed by G. Siu (California Institute of Technology, unpublished observations). Cosmid clone 17.1 was isolated from a BALB/cJ liver DNA cosmid library constructed in this study. All dm1 cosmid clones were isolated from the B10.D2-H-2^{dm1} liver DNA cosmid library constructed in this study.

Restriction Mapping of Cosmid Clones. Restriction map analyses were performed using a combination of the standard double-digestion method and a new, rapid method. The new method is a modification of the method by Smith and Birnstiel (29) and by Rackwitz et al. (30). Briefly, two 40-mer oligonucleotides were synthesized corresponding to sequences flanking each side of the Sal I recognition site in pTL5. Sal I cleaves once in the vector but very rarely cuts within the insert. The cosmid DNA were linearized with Sal I, partially digested with the restriction enzymes used for mapping, electrophoresed on agarose gel, blotted onto NC, and hybridized to the radioactively-labeled oligonucleotides. The two end-specific oligonucleotides will each hybridize to a set of partial fragments, all with a common terminus but of different lengths, forming a ladder of bands that can be detected by autoradiography. The positions of all the restriction sites relative to the Sal I site can therefore be easily deduced. Detailed procedures will be published separately.

DNA Sequencing. The 2.4 kb Bam HI fragment containing the 5′ half of the D^{dm1} gene was subcloned into the Bam HI site of M13mp8. This subclone was mapped with the restriction enzymes Sma I, Pst I, and Xba I. From the cloned 2.4 kb fragment, a 300 base pair (bp) Sma I/Pst I fragment and a 500 bp Pst I/Xba I fragment were separately subcloned into the corresponding sites in both M13mp10 and mp11; a 300 bp and a 50 bp Sma I fragment were subcloned into M13mp8 Sma I site. DNA sequence analyses were carried out by the chain-terminating method of Sanger et al. (31).

Transfection and Radioimmunoassay. Transfection of mouse L cells with cosmid clone 40.2 and cell-surface radioimmunoassay were performed as described previously (27, 32).

Results

DNA Sequences 3′ to the D^d Gene and 5′ to the Ld Gene are Deleted in the dm1 Mouse. Because both D^d and L^d antigens are affected in the dm1 mutant, we first investigated whether the mutation affected the structural genes. Since the class I gene family contains 33 crosshybridizing members (26, 33), it is difficult to use coding sequences as probes to distinguish among the L^d, D^d, and other class I genes. Instead, several low copy-number sequences were subcloned from the more divergent regions flanking the L^d and D^d genes (Fig. 1). These were then used as hybridization probes in genomic Southern blot analyses to detect differences between the dm1 mutant and its parental strain B10.D2. BALB/c DNA was also included in the study because the D^d and L^d cosmid clones were generated from this strain (26). The BALB/c H-2 complex appears to be identical to the B10.D2 H-2 complex.

Probe 59.2A, located 2–4 kb 3′ of the L^d gene (Fig. 1), detects four Bam HI bands in BALB/c and B10.D2 liver DNA (Fig. 2). These four fragments have

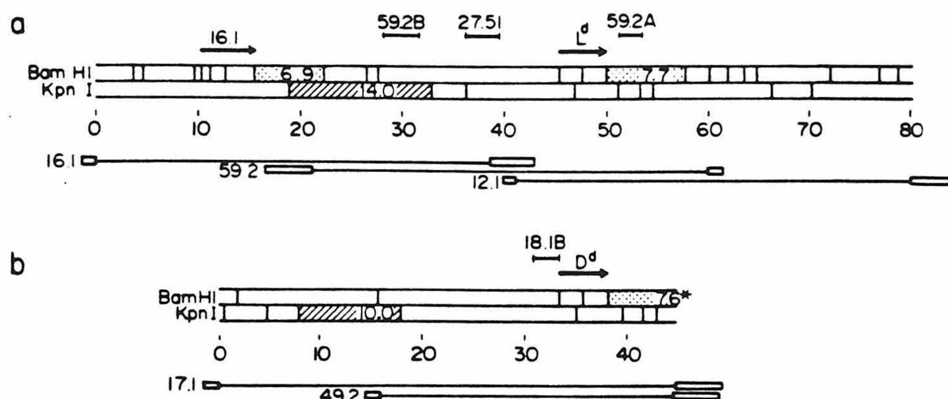


FIGURE 1. Restriction map of the L^d and D^d clusters and location of hybridization probes. Restriction map of the cosmid clusters containing the L^d gene (*a*) and the D^d gene (*b*). The regions contained in each cosmid clone are shown below the map with open boxes at the ends denoting the vector arms split at the single Sal I site. The class I genes are denoted by a heavy bar with arrow indicating the transcription orientation. Location of the hybridization probes are shown by a bar above the restriction map. The Bam HI fragments that hybridize to probe 59.2A are dotted, the Kpn I fragments that hybridize to probe 59.2B are crosshatched. The sizes of these fragments are shown in kbp. Bam HI fragment (*) is not complete in clone 17.1, but a cosmid clone, 2.20, overlapping with clone 17.1 (M. Steinmetz, Basel Institute for Immunology, personal communication), contains the complete fragment, which is 7.6 kb (data not shown).

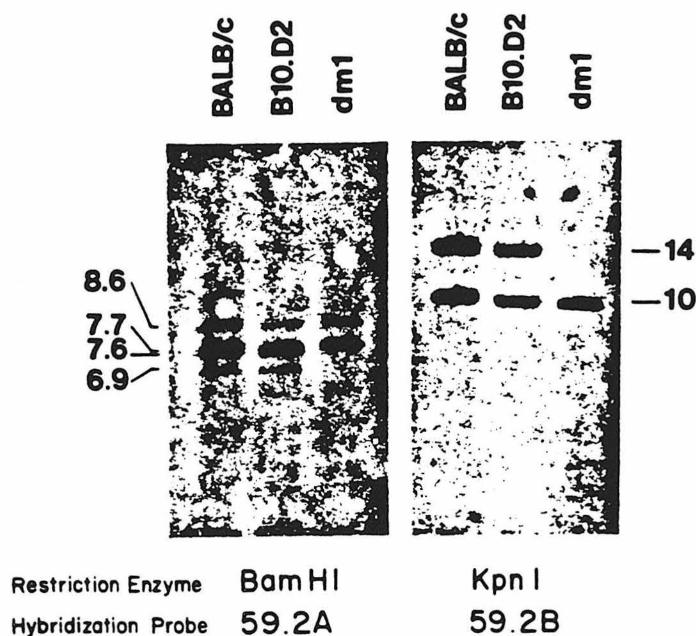


FIGURE 2. Southern blot analyses of genomic DNA showing deletion in the dm1 mutant. 10 μg of liver DNA from the mouse strains BALB/c, B10.D2, and B10.D2-H-2^{dm1} were digested with Bam HI or Kpn I, electrophoresed on a 0.6% agarose gel, blotted, and hybridized to radioactively-labeled probes 59.2A or 59.2B. All of the hybridizing restriction fragments have been identified in BALB/c cosmid clones. Their locations are shown in Fig. 1.

been identified in cosmid clones (Fig. 1). They all lie immediately 3′ to an H-2 class I gene. The 8.6 kb fragment is associated with gene 17.1, which is ~19 kb 3′ of the K^d gene (data not shown). The 7.7 kb fragment, from which the probe was generated, is associated with the L^d gene. The 7.6 kb fragment is associated with the D^d gene (see Fig. 1 legend). The 6.9 kb fragment is associated with gene 16.1, which is 5′ to the L^d gene. Southern blot analysis of liver DNA with probe 59.2A shows that the 7.6 kb and 6.9 kb Bam HI fragments are missing in the dm1 mouse, indicating that the DNA in the regions 5′ to the L^d gene and 3′ to the D^d gene have been deleted.

Probe 59.2B, located 13 kb 5′ of the L^d gene (Fig. 1), detects 14 and 10 kb Kpn I bands in BALB/c and B10.D2 liver DNA (Figs. 1 and 2). The 14 kb band corresponds to the fragment in clone 59.2 from which the probe was generated. This fragment is not present in dm1, indicating the region 5′ to the L^d gene has been deleted. Probe pH2^{d}27.51, located 5 kb 5′ of the L^d gene (Fig. 1), also fails to detect its corresponding fragment in dm1 (data not shown), indicating that the 5′ deletion extends within 5 kb of the L^d gene. Hybridization with probe 18.1B, located immediately 5′ to the D^d gene (Fig. 1), shows that the corresponding fragment is present in dm1 (data not shown).

These results indicate a deletion of the region 5′ to the L^d gene, and the region 3′ to the D^d gene, while the regions 3′ to the L^d gene and 5′ to the D^d gene are unaltered.

The dm1 Mutant Gene Is a D/L Hybrid Gene. To study the extent of the deletion, a cosmid library was constructed from dm1 liver DNA in order to clone the altered region. The library was screened using probes 59.2A and 59.2B, and 14 positive clones were found. By comparing single digests using several restriction enzymes, the clones were shown to be overlapping, and their order was determined. Three clones (38.1, 40.2, 7.2) that cover the whole region were then mapped in detail (Fig. 3, *A* and *B*) using a rapid restriction mapping method in combination with the conventional double-digest method. The cosmid cluster spans 90 kb and contains one class I gene. When the restriction map is compared with those of the L^d and D^d cosmid clusters, the region upstream of this gene is identical to that of D^d, while the region downstream of this gene is identical to that of the L^d gene (Fig. 3*C*). A 15 kb region including the *dm1* class I gene is indistinguishable by restriction map (Fig. 3 *C*) with the corresponding region of the L^d and the D^d gene clusters. These data suggest that the mutation resulted from a deletion removing the DNA between the D^d and L^d genes, possibly forming a hybrid gene.

To further define the region where the deletion occurred, restriction site polymorphisms around the L^d and D^d genes were identified in order to distinguish the two loci. The Sma I site in the second intron of the D^d gene was resolved into a doublet spaced 52 bp apart (28), while only one Sma I site is found in the corresponding region of the L^d gene. The additional Sma I site is present in the dm1 gene (data not shown), indicating that the D^d locus contributed the 5′ region of the dm1 gene extending at least 3′ to the position of the Sma I site. Studies with cytotoxic T lymphocytes (CTL) specific to amino acid residues 152–156 of the L^d antigen suggested that the 3′ end of the gene, extending 5′ to at least the codon for amino acid residue 152 is donated by the L^d gene (24).

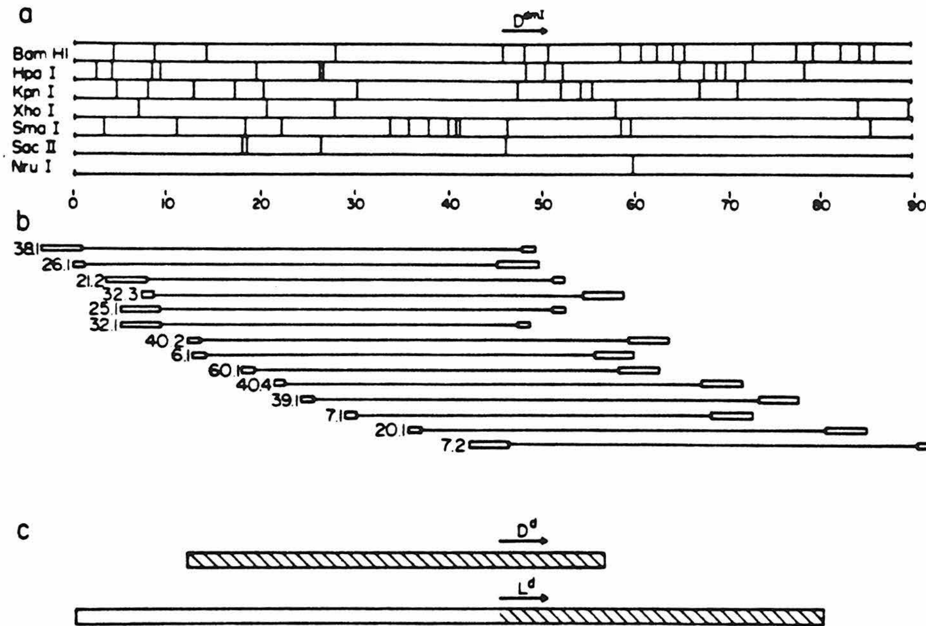


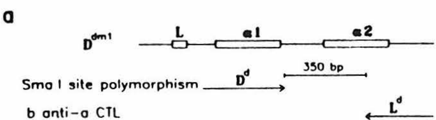
FIGURE 3. Restriction map for the 14 overlapping dm1 cosmid clones. (*a*) The composite restriction map is shown. Sal I and Cla I sites were not found. The mapping was done by a combination of double digestion and a rapid method as described in Materials and Methods. Lengths given in kb. (*b*) The regions contained in each cosmid clone are shown in lines with open boxes at the ends denoting the vector arms split at the single Sal I site. (*c*) Comparison of restriction map of the D^d and L^d with the dm1 cosmid clusters. Hatched region denotes indistinguishable restriction map to the dm1 clones.

Combination of the data from the Sma I site polymorphism and the CTL studies suggests that the *dm1* gene is a hybrid gene formed by a fusion between the 5′ portion of the D^d gene and the 3′ portion of the L^d gene (Fig. 4*A*). The recombination point apparently lies within a region of ~350 bp. This D/L hybrid gene is called the D^{dm1} gene because it is presumably transcribed from the D^d promoter.

DNA sequence analysis was carried out in this region of the D^{dm1} gene and compared against D^d and L^d sequences. The recombination point occurs in exon 3, which encodes the $\alpha 2$ region. As shown in Fig. 4*B*, the sequence from the beginning of exon 3 to position 827 (following the numbering in the D^d sequence) is identical to the D^d sequence (28), while the 3′ portion of the exon beginning at nucleotide 927 is identical to the L^d sequence (34). The 99 bp of sequence between these points is identical in the L^d and D^d genes. The sequence analysis clearly shows that the recombination event occurred within this 99 bp stretch. The resulting D/L hybrid antigen has the $\alpha 1$ and two-thirds of the $\alpha 2$ region identical to the D^d antigen, and two-thirds of the $\alpha 2$ region and the carboxyl half of the molecule identical to the L^d antigen (Fig. 4*C*).

The D^{dm1} Gene Is Expressed in Transfected Fibroblasts and Correlates with the Mutant Phenotype. The D^{dm1} gene was transfected into mouse L cells. The transformant line K31-25 was tested with a panel of mAb by cell-surface radioimmunoassay (Fig. 5). These mAb recognize specific regions of different H-2 class

a



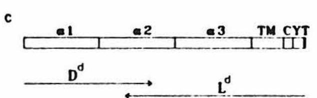
b

```
D^dm1  91                                          lySerHisThrLeuGlnTrpMetAlaGlyCysAspValGluSerAspGlyArgLeuLeuArgGlyTyrTrpGlnPheAla 117
          GGGCCGGGGCGGGTGAGCGGGGCTGACCGGGGTCCCGCAGGCTCTCACACACTCCAGTGGATGGCTGGCTGTGACGTGGAGTCGGACGGCGCCTCCTCCGCGGGTACTGGCAGTTCGCC
D^d       ------------------------------------------------------------------------------------------------------------------------
                                                  Thr                   Tyr            Gly                           Glu
L^d       ----------------------------------------A---------------------TAC-------------G-----------------------------GA----------

D^dm1 118 TyrAspGlyCysAspTyrIleAlaLeuAsnGluAspLeuLysThrTrpThrAlaAlaAspMetAlaAlaGlnIleThrArgArgLysTrpGluGlnAlaGlyAlaAlaGluTyrTyrArg 157
          TACGACGGCTGCGATTACATCGCCCTGAACGAAGACCTGAAAACGTGGACGGCGGCGGACATGGCGGCGCAGATCACCCGACGCAAGTGGGAGCAGGCTGGTGCTGCAGAGTATTACAGG
                                                                                                                    ArgAsp
D^d       ------------------------------------------------------------------------------------------------------------AGAG--C--
                  Arg
L^d       --------C-T-------------------------------------------------------------------------------------------------------------

D^dm1 158 AlaTyrLeuGluGlyGluCysValGluTrpLeuHisArgTyrLeuLysAsnGlyAsnAlaThrLeuLeuArgThrG                                            182
          GCCTACCTGGAGGGCGAGTGCGTGGAGTGGCTCCACAGATACCTGAAGAACGGGAACGCGACGCTGCTGCGCACAGGTGCAGGGCCGCGGGCAG
                                      Arg
D^d       ------------------------------G-----------------------T---------------------------------------
L^d       ----------------------------------------------------------------------------------------------
```

c



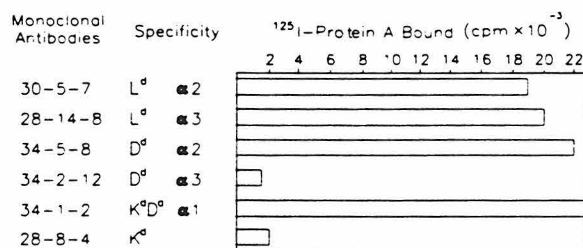


FIGURE 5. Cell-surface radioimmunoassay of the transformant K31-25. Cosmid clone 40.2 was used to transfect mouse L cells. The transformant line K31-25 was tested with a panel of mAb. The regions that the mAb recognize are indicated (35, 36).

I antigens (35, 36). The transformant line reacts with mAb that recognize either the D^d $\alpha 1$ region or the L^d $\alpha 3$ region. These results confirm that the cloned gene encodes a hybrid class I antigen with its $\alpha 1$ region derived from the D^d gene and the $\alpha 3$ region derived from the L^d gene. mAb 30-5-7 and 34-5-8, specific for the $\alpha 2$ region of L^d and D^d antigens, respectively, both react with the transformant. Taken together with the DNA analyses, these results further assign the specificity of mAb 30-5-7 to the carboxyl-terminal third of the L^d $\alpha 2$ region, and the specificity of mAb 34-5-8 to the amino-terminal third of the D^d $\alpha 2$ region.

The pattern of reactivity of the transformant with the panel of mAb is identical to that of dm1 spleen cells (24, 46). Preliminary results from two-dimensional gel analysis of the immunoprecipitated D^{dm1} antigen from the transformant also indicate it is identical to that from the dm1 spleen cells (M. McMillan, University of Southern California, Los Angeles, CA, personal communication). These data unequivocally show that the D^{dm1} gene codes, for the mutant class I antigen.

Discussion

Possible Mechanisms for Generating the Fusion Gene. We demonstrate here that the cloned D^{dm1} gene is a hybrid of the 5′ end of the D^d gene and the 3′ end of the L^d gene, with the region in between deleted. This gene fusion can presumably occur by either intrachromosomal deletion or interchromosomal unequal crossover (Fig. 6). In either case, it is most likely the result of homologous recombination, since the region where the recombination occurred is a 99 bp stretch of sequences identical between the D^d and the L^d genes. The reciprocal product (the L/D hybrid gene) in the recombination event presumably segregated from

FIGURE 4. The *dm1* gene is a D/L hybrid gene. (*a*) The left boundary of the 350 bp region is defined by the polymorphic Sma I site at 12 bp 3′ of exon II (encoding the $\alpha 1$ region). The right boundary is at amino acid residue 151, as defined by the b anti-a CTL. Recombination occurred within this 350 bp region. (*b*) The 2.4 kb Bam HI fragment containing the 5′ half of the dm1 gene was subcloned and mapped. ~1.2 kb has been sequenced from M13 subclones from the 2.4 kb Bam HI fragment. DNA sequence is only shown for the region immediately containing the third exon, where the recombination occurred. The dm1 sequence is aligned with the DNA sequence of L^d (34) and of D^d (28). Identity is indicated by a dash. Translated protein sequence of the D^{dm1} gene is shown above its DNA sequence. Differences in L^d and D^d are indicated. (*c*) The predicted D^{dm1} protein structure is shown. The $\alpha 1$ region and two-thirds of the $\alpha 2$ region are identical with the D^d antigen. Two-thirds of the $\alpha 2$ region, the $\alpha 3$ region, the transmembrane (*TM*) region, and the cytoplasmic regions (*CYT*) are identical to the L^d antigen.

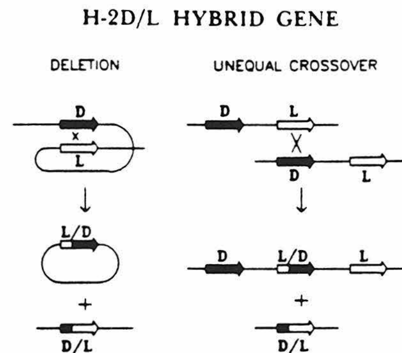


FIGURE 6. Two models for the occurrence of the dm1 hybrid gene.

the D/L hybrid gene in meiosis and was not passed on to progeny. Mice bearing such a reciprocal mutation have not been described.

In chromosomal walking experiments (Y. Sun, unpublished observations), the L^d gene has been linked to the Qa-2,3 cosmid cluster, with the transcription direction of L^d towards the *Qa-2,3* gene. This establishes that the D^d gene is located upstream of the L^d gene. This is consistent with the gene order reported in a recent study (47) using a recombinant between the H-2D and H-2L genes. In the above models, it is assumed that the D^d and the L^d genes have the same transcriptional orientation. In this configuration, only a single recombination event is required to create the mutation. Since the D^d gene has not been physically linked to the distal L^d gene nor to the proximal S region, which contains the class III genes, it remains to be shown whether the D^d or L^d genes do have the same transcriptional orientation.

Implications for Evolution of Multigene Families. Other examples of gene fusion have been found. Two hemoglobin variants, Lepore-Boston and Kenya, have been characterized by restriction site analyses (48, 49) of genomic DNA, and were shown to be due to gene fusion. Based on restriction map comparison, the BALB/c *Q7* gene has been suggested (50) to be a fusion between the equivalents of genes *Q8* and *Q9* of the B10 mouse. The *Q3* gene of the BALB/c mouse has also been suggested to have resulted from a gene fusion (S. Hunt, California Institute of Technology, personal communication).

Homologous but unequal recombination, as suggested by these examples of gene fusion, can serve to expand or contract the number of genes in a multigene family. Contraction and expansion of the class I gene repertoire is supported by many studies (2, 50–52) comparing different mouse inbred strains and different *Mus* species. Our study of the dm1 mutation is the first direct evidence of the contraction of the size of the H-2 class I gene family. Preliminary results show that the dm2 mutation (L^d loss mutant) also involves a DNA deletion, possibly including the L^d gene (Y. Sun, unpublished data), providing another example of this contraction process. Recombination at regions within the structural genes not only changes the size of the gene family but also can immediately create new (hybrid) members in the gene family, hence it serves as a mechanism for directly generating polymorphism. The dm1 mutant also provides an example of this process.

One of the most interesting and perplexing aspects of the H-2 transplantation

antigens is their extremely high polymorphism (37). Gene conversion and unequal crossover have been suggested as mechanisms for generating the polymorphism (2). Both mechanisms probably begin with the same molecular configuration of pairing between homologous sequences. However, a notable difference of their operation is noted in the mutants of two H-2 haplotypes. Of the 21 H-2^b mutants, 14 have been analyzed biochemically, and the alterations have been localized to 1–4 amino acids (7, 38). Gene conversion of short stretches of DNA sequence has been suggested as the mechanism generating these mutations, based on the following observations: the altered amino acids are clustered in each mutant; some of the clustered changes occurred in several independently derived mutants; and in several cases, a potential donor class I gene can be identified (38–41 and J. Geliebter, Albert Einstein College of Medicine, New York, NY, personal communication). On the other hand, mutants in the H-2^d have more extensive alterations. The dm1 and dm2 mutations involve DNA deletion. The dm5, an H-$2K^d$ mutant, can be shown by Southern blot analysis to have a significantly altered pattern of class I genes (R. Goodenow, unpublished data). The dm6 mutant fails to express the D^d antigen (19), possibly also as a consequence of DNA deletion. The basis for this predominance of gene conversions in the H-2^b haplotype and recombination in the H-2^d haplotype is not known.

Functional Dissection of the Transplantation Antigens. Using hybrid class I genes constructed by exon-shuffling methods, it has been possible to determine which regions of the H-2 class I antigen are important for the interactions with antibody and CTL (35, 36, 42–45). The dm1 mutation provides a natural hybrid gene. Whereas the hybrid genes artifically constructed to date have exchanged complete axons, the D^{dm1} gene has a hybrid $\alpha 2$ exon, allowing a finer functional dissection within a single external region. For example, as shown in the previous section, the position responsible for the specificity by mAb 30-5-7 and 34-5-8 can be narrowed down to only one-third of the $\alpha 2$ region.

Studies with constructed hybrid class I genes have shown that the $\alpha 1$ and $\alpha 2$, but not $\alpha 3$, regions are important in determining the specificity of recognition by alloreactive or antiviral CTL (43, 44). It has also been shown (45, 53) that the $\alpha 1$ and $\alpha 2$ regions cannot be independently recognized by CTL, suggesting that CTL recognition requires both regions. CTL from lymphocytic choriomeningitis virus–infected dm1 mouse are restricted by the D^{dm1} antigen, but not by the K^d, D^d, or L^d antigens (54). Since the D^{dm1} and the D^d antigen share the $\alpha 1$ region and two-thirds of the $\alpha 2$ region, it can be concluded that this shared region is not sufficient, and the last one-third of the $\alpha 2$ region is important in determining the specificity of CTL recognition. The last one-third of the α_2 region, however, is not sufficient because the L^d antigen shares the same sequence, but does not restrict dm1 CTL. Anti-L^d alloreactive CTL do not lyse dm2 cells (D^d-positive, L^d-negative) as expected, but show strong, although incomplete, crossreactivity for dm1 target cells (55), again showing that the last one-third of $\alpha 2$ region is important in CTL recognition. These data are consistent with conclusions cited above that the CTL recognize determinants that are formed by or are sensitive to interactions between the α_1 and α_2 regions, which have been suggested by x-ray crystallography studies (56) to fold and jointly form a domain.

None of the Additional H-2^d Class I Antigens Is Found in the dm1 Mouse. Several additional H-2^d class I antigens, including M^d, R^d, $L2^d$, identified either by sequential immunoprecipitation or cocapping experiments, have been mapped (3) to the *H-$2D^d$* region. The molecular basis of these additional class I antigens is not clear. They may be encoded by genes other than L^d and D^d, or be derived from the L^d or D^d gene via an alternative RNA splicing pathway or with different posttranslational modification (27). None of these additional antigens can be detected on the dm1 spleen cell nor in the cytoplasm (24, 25, 57). This suggests that the genes encoding these antigens have been deleted, or that the regulatory mechanisms for their expression are not functional in the dm1 mutant.

From our study of the dm1 mutation, two points can be made regarding these proposed mechanisms. First, putative alternative splice donor and acceptor sites have been identified in the second exon (encoding the $\alpha 1$ region) of the D^d gene (28). These sites are also present in the D^{dm1} gene. Since neither M^d nor R^d antigens were found in the dm1 mouse, they are probably not derived from the D^d gene by alternative RNA splicing. A study with the L^d transformants (58) also suggested that the R^d antigen is not derived from the L^d gene. Second, gene 16.1 (see Fig. 1) could encode the R^d antigen. The R^d antigen is probably a secreted or cytoplasmic molecule because it can be detected in cell lysate (46), but not on the cell surface (59). It also has a lower molecular mass (58), consistent with lack of a transmembrane region. Mouse L cells transfected with gene 16.1 do not express any new cell surface molecules recognized by the available H-2^d antisera and mAb. It may be a pseudogene, or it may encode a secreted or cytoplasmic antigen, possibly the R^d antigen. Consistent with the latter hypothesis is the absence of the R^d antigen in dm1 mouse, in which gene 16.1 has been deleted. Its location upstream of the H-$2L^d$ gene is also consistent with the gene order of D-R-L (C. David, Mayo Medical School, Rochester, MN, personal communication). Analysis of cell lysate from L cells transformed with this gene will readily provide a test for this hypothesis.

Summary

The H-2^{dm1} mutation is unique among all described H-2 mutations in that two transplantation antigens, the H-$2D^d$ and the H-$2L^d$, are affected. Here, we show that the mutant gene, D^{dm1}, is formed by fusion of the 5′ part of the D^d gene and the 3′ part of the L^d gene, with the region in between deleted. The recombination junction is located in the third exon, which encodes the $\alpha 2$ region of the protein. When the hybrid gene is transfected into mouse L cells, serological and biochemical analyses indicate the D^{dm1} antigen expressed in the transformant line is identical to the mutant molecule in dm1 spleen cells. These results demonstrate that the D/L hybrid gene is most likely responsible for the dm1 mutant phenotype.

We thank C. Zupke and J. Luo for technical assistance, and C. Katz for preparation of the manuscript. The authors are grateful to G. Siu, D. Fisher, P. Lonai, I. Stroynowski, and U. Landegren for helpful comments on the manuscript.

Received for publication 30 July 1985.

References

1. Zinkernagel, R. M., and P. C. Doherty. 1979. MHC-restricted cytotoxic T cells: Studies on the biological role of polymorphic major transplantation antigens determining T cell restriction specificity, function and responsiveness. *Adv. Immunol.* 27:51.
2. Hood, L., M. Steinmetz, and B. Malissen. 1983. Genes of the major histocompatibility complex of the mouse. *Annu. Rev. Immunol.* 1:529.
3. Hansen, T. H., K. Ozato, and D. H. Sachs. 1983. Heterogeneity of H-2D region associated genes and gene products. *Adv. Immunol.* 34:39.
4. Démant, P., and D. Iványi. 1981. Further molecular complexities of H-2K- and D-region antigens. *Nature (Lond.).* 290:146.
5. Kohn, H., J. Klein, R. W. Melvold, S. G. Nathenson, D. Pious, and D. Shreffler. 1978. The first H-2 mutant workshop. *Immunogenetics.* 7:279.
6. Klein, J. 1978. H-2 mutations: their genetics and effect on immune functions. *Adv. Immunology.* 26:55.
7. Nairn, R., K. Yamaga, and S. G. Nathenson. 1980. Biochemistry of the gene products from murine MHC mutants. *Annu. Rev. Genet.* 14:241.
8. Melvold, R. W., H. I. Kohn, and G. R. Dunn. 1982. History and genealogy of the H-$2K^{b}$ mutants from the C57BL/6Kh colony. *Immunogenetics.* 15:177.
9. Melvold, R. W., P. M. Stuart, A. E. Busker, and B. Beck-Mair. 1983. A new loss mutant in the H-2^{d} haplotype. *Transplant. Proc.* 15:2045.
10. Egorov, O. S., and I. K. Egorov. 1984. H-2^{bm23}, a new K^{b} mutant similar to, but not identical with H-2^{bm3}. *Immunogenetics.* 20:83.
11. McKenzie, I. F. C., T. Pang, and R. V. Blanden. 1977. The use of H-2 mutants as models for the study of T cell activation. *Immunol. Rev.* 35:181.
12. Ewald, S. J., J. Klein, and L. E. Hood. 1979. Peptide map analysis of mutant transplantation antigens. *Immunogenetics.* 8:551.
13. Egorov, I. K. 1967. A mutation of the histocompatibility-2 locus of the mouse. *Genetika.* 3:136.
14. Klein, J., and I. K. Egorov. 1973. Graft-vs.-host reaction with an H-2 mutant. *J. Immunol.* 111:976.
15. Klein, J., M. Hauptfeld, R. Geib, and C. Hammerberg. 1976. Immunogenetic analysis of H-2 mutations. VI. Serological analysis of mutations H-2da, H-2^{fa}, and H-2^{ka1}. *Transplantation (Baltimore).* 22:572.
16. Dishkant, I. P., A. A. Vedernikov, and I. K. Egorov. 1973. Study of H-2 mutations in mice. III. Serological analysis of the mutation 504 and its derived recombinant H-2 haplotypes. *Genetika.* 9:82.
17. Shumova, T. E., D. P. Kryshkina, and I. K. Egorov. 1972. Study of H-2 mutations of mice: Analysis of mutation 504 by the method of F_1 hybrids. *Genetika.* 8:171.
18. Vedernikov, A. A., and I. K. Egorov. 1973. Study of H-2 mutations in mice. II. Recombination analysis of the mutation 504. *Genetika.* 9:60.
19. McKenzie, I. F. C., G. M. Morgan, R. W. Melvold, and H. I. Kohn. 1977. BALB/c-H-2^{db}: a new H-2 mutant in BALB/cKh that identifies a locus associated with the D region. *Immunogenetics.* 4:333.
20. Morgan, G. M., I. F. C. McKenzie, and R. W. Melvold. 1978. The relationship between the H-2 locus mutations of H-2da and H-2^{db} in the mouse. *Immunogenetics.* 7:247.
21. Iványi, D., and P. Démant. 1978. Complex genetic effect of B10.D2 (M504) (H-2^{dm1}) mutation. *Immunogenetics.* 8:539.
22. Brown, J. L., R. Nairn, and S. G. Nathenson. 1978. Structural differences between

the mouse H-2D products of the mutant B10.D2.M504 (H-2da) and the parental nonmutant strain B10.D2(H-2^d). *J. Immunol.* 120:726.
23. Wilson, P. H., R. Nairn, S. G. Nathenson, and D. W. Sears. 1982. Unexpected complexity of the dm1 mutation revealed in the structure of three H-2D/L-related antigens. *Immunogenetics.* 15:225.
24. Burnside, S. S., P. Hunt, K. Ozato, and D. W. Sears. 1984. A molecular hybrid of the H-$2D^d$ and H-$2L^d$ genes expressed in the dm1 mutant. *Proc. Natl. Acad. Sci. USA.* 81:5204.
25. Hunt, P., S. S. Burnside, and D. W. Sears. 1985. Evidence of only one H-2D/L-related glycoprotein in H-2^{dm1} mutant cells. *J. Immunogenet. (Oxf.).* 12:33.
25a. Winoto, A., M. Steinmetz, and L. Hood. 1983. Genetic mapping in the major histocompatibility complex by restriction enzyme site polymorphism: Most mouse class I genes map to the *Tla* complex. *Proc. Natl. Acad. Sci. USA.* 80:3425.
26. Steinmetz, M., A. Winoto, K. Minard, and L. Hood. 1982. Clusters of genes encoding mouse transplantation antigens. *Cell.* 28:489.
27. Goodenow, R. S., M. McMillan, M. Nicolson, B. T. Sher, K. Eakle, N. Davidson, and L. Hood. 1982. Identification of the class I genes of the mouse major histocompatibility complex by DNA-mediated gene transfer. *Nature (Lond.).* 300:231.
27. Hu, N.-T., and J. Messing. 1982. The making of strand-specific M13 probes. *Gene (Amst.).* 17:271.
28. Sher, B. T., R. Nairn, J. E. Coligan, and L. Hood. DNA sequence of the mouse H-$2D^d$ transplantation gene. *Proc. Natl. Acad. Sci. USA.* 82:1175.
29. Smith, H., and M. L. Birnstiel. 1976. A simple method for DNA restriction site mapping. *Nucleic Acids Res.* 3:2387.
30. Rackwitz, H.-R., G. Zehetner, A.-M. Frischauf, and H. Lehrach. 1984. Rapid restriction mapping of DNA cloned in lambda phage vectors. *Gene (Amst.).* 30:195.
31. Sanger, F., S. Hicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463.
32. Goodenow, R. S., M. McMillan, A. Orn, M. Nicolson, N. Davidson, J. A. Frelinger, and L. Hood. 1982. Identification of a BALB/c H-$2L^d$ gene by DNA-mediated gene transfer. *Science (Wash. DC).* 215:677.
33. Fisher, D. A., S. W. Hunt, and L. Hood. 1985. The structure of a gene encoding a murine thymus leukemia antigen and the organization of *Tla* genes in the BALB/c mouse. *J. Exp. Med.* 162:528.
34. Moore, K. W., B. T. Sher, Y. H. Sun, K. A. Eakle, and L. Hood. 1982. DNA sequence of a gene encoding a BALB/c mouse L^d transplantation antigen. *Science (Wash. DC).* 215:679.
35. Evans, G. A., D. H. Margulies, B. Shykind, J. G. Seidman, and K. Ozato. 1982. Exon-shuffling: mapping polymorphic determinants on hybrid mouse transplantation antigens. *Nature (Lond.).* 300:755.
36. Murre, C., E. Choi, J. Weis, J. G. Seidman, K. Ozato, L. Liu, S. J. Burakoff, and C. S. Reiss. 1984. Dissection of serological and cytolytic T lymphocyte epitopes on murine major histocompatibility antigens by a recombinant H-2 gene separating the first two external domains. *J. Exp. Med.* 160:167.
37. Klein, J. 1979. The major histocompatibility complex of the mouse. *Science (Wash. DC).* 203:516.
38. Pease, L. R., D. H. Schulze, G. M. Pfaffenbach, and S. G. Nathenson. 1983. Spontaneous H-2 mutants provide evidence that a copy mechanism analogous to gene conversion generates polymorphism in the major histocompatibility complex. *Proc. Natl. Acad. Sci. USA.* 80:242.
39. Mellor, A. L., E. H. Weiss, K. Ramachandran, and R. A. Flavell. 1983. A potential

donor gene for the *bm1* gene conversion event in the C57BL mouse. *Nature (Lond.).* 306:782.

40. Schulze, D. H., L. E. Pease, S. S. Grier, A. A. Reyes, L. A. Sarmiento, R. B. Wallace, and S. G. Nathenson. 1983. Comparison of the cloned H-2K^{bm1} variant gene with the H-2K^{b} gene shows a cluster of seven nucleotide differences. *Proc. Natl. Acad. Sci. USA.* 80:2007.
41. Weiss, E. H., A. Mellor, L. Golden, K. Fahrner, E. Simpson, J. Hurst, and R. A. Flavell. 1983. The structure of a mutant H-2 gene suggests that the generation of polymorphism in H-2 genes may occur by gene conversion–like events. *Nature (Lond.).* 301:671.
42. Ozato, L., G. A. Evans, B. Shykind, D. Margulies, and J. G. Seidman. 1983. Hybrid H-2 histocompatibility gene products assign domains recognized by alloreactive T cells. *Proc. Natl. Acad. Sci. USA.* 80:2040.
43. Reiss, C., G. A. Evans, D. H. Margulies, J. G. Seidman, and S. J. Burakoff. 1983. Allospecific and virus-specific cytolytic T lymphocytes are restricted to the N or C1 domain of H-2 antigens expressed on L cells after DNA-mediated gene transfer. *Proc. Natl. Acad. Sci. USA.* 80:2709.
44. Stroynowski, I., A. Orn, R. S. Goodenow, M. McMillan, J. Forman, P. R. Brayton, J. Frelinger, and L. Hood. 1984. Cytotoxic T lymphocytes recognize determinants on the BALB/c-H-2L^{d} molecule controlled by $\alpha1$ and $\alpha2$ but not $\alpha3$ external domains. *Immunogenetics.* 20:141.
45. Allen, H., D. Wraith, P. Pala, B. Askonas, and R. A. Flavell. 1984. Domain interactions of H-2 class I antigens alter cytotoxic T-cell recognition sites. *Nature (Lond.).* 309:279.
46. Hansen, T. H., K. Ozato, M. R. Melino, J. E. Coligan, T. J. Kindt, J. J. Jandinski, and D. H. Sachs. 1981. Immunochemical evidence in two haplotypes for at least three D region–encoded molecules, D, L, and R. *J. Immunol.* 126:1713.
47. Savarirayan, S., W. P. Lafuse, and C. S. David. 1985. Recombination between H-2D and H-2L genes: Identification, characterization, and gene order. *Transplant. Proc.* 17:702.
48. Flavell, R. A., J. M. Kooter, E. DeBoer, P. F. R. Little, and R. Williamson. 1978. Analysis of the β–δ globin gene loci in normal and Hb Lepore DNA: Direct determination of gene linkage and intergene distance. *Cell.* 15:25.
49. Ojwang, O. J., T. Nakatsuji, M. B. Gardiner, A. L. Reese, J. G. Gilman, and T. H. J. Huisman. 1983. Gene deletion as the molecular basis for the Kenya-G$^{\gamma}$-HPFH condition. *Hemoglobin.* 7:115.
50. Weiss, E. H., L. Golden, K. Fahrner, A. L. Mellon, J. J. Devlin, H. Bullman, H. Tiddens, H. Bud, and R. A. Flavell. 1984. Organization of the class I gene family in the major histocompatibility complex of the C57BL/10 mouse. *Nature (Lond.).* 310:650.
51. Steinmetz, M., M. Malissen, L. Hood, A. Orn, R. A. Maki, G. R. Dastoornikoo, D. Stephan, E. Gibb, and R. Rananink. 1984. Tracts of high or low sequence divergence in the mouse major histocompatibility complex. *EMBO (Eur. Mol. Biol. Organ.) J.* 3:2995.
52. Rogers, M. J., R. N. Germain, J. Hare, E. Long, and D. S. Singer. 1984. Comparison of MHC genes among distantly related members of the genus *Mus. J. Immunol.* 134:630.
53. Stroynowski, I., S. Clark, L. A. Henderson, L. Hood, M. McMillan, and J. Forman. 1985. Interaction of $\alpha1$ with $\alpha2$ region in class I MHC proteins contributes determinants recognized by antibodies and cytotoxic T cells. *J. Immunol.* 135:2160.
54. Zinkernagel, R. M., and J. Klein. 1977. H-2-associated specificity of virus-immune

cytotoxic T cells from H-2 mutant and wild-type mice: M523 (H-$2K^{ka}$) and M505 (H-$2K^{bd}$) do, M504 (H-$2K^{da}$) and M506 (H-$2K^{fa}$) do not crossreact with wild-type H-2K or H-2D. *Immunogenetics.* 4:581.
55. Melief, C. J. M., L. P. de Waal, M. Y. van der Menlen, P. Iványi, and R. W. Melvold. 1981. Fine specificity of cytotoxic T lymphocytes directed against H-$2L^d$. *Immunogenetics.* 12:75.
56. Bjökman, P. 1984. Ph.D. thesis. Harvard University, Cambridge, MA.
57. Iványi, D., and P. Démant. 1982. Molecular heterogeneity of D-end products detected by anti-H-2.28 sera. II. B10.D2 (M504) (H-2^{dm1}) mutant fails to express one of the two H-2.4^-, 28^+ D^d region molecules. *Immunogenetics.* 15:467.
58. Nichols, E. A., T. Krakaner, and T. H. Hansen: 1983. Two-dimensional gel comparison of murine H-2D region–associated antigens of different H-2 haplotypes. *J. Immunol.* 131:2440.
59. Iványi, D., and P. Démant. 1983. Capping experiments fail to reveal H-$2R^d$ molecules in D^d region. *Transplant. Proc.* 15:2039.

CHAPTER TWO

ORGANIZATION AND EVOLUTION OF *D* REGION CLASS I GENES IN THE MOUSE MAJOR HISTOCOMPATIBILITY COMPLEX

Submitted for publication to the *Journal of Experimental Medicine*

ORGANIZATION AND EVOLUTION OF *D* REGION CLASS I GENES IN THE MOUSE MAJOR HISTOCOMPATIBILITY COMPLEX

By Danielle Stephan*, Henry Sun++, Kirsten Fischer Lindahl*, Evi Meyer§, Günter Hämmerling§, Leroy Hood++, and Michael Steinmetz*

From the:

*Basel Institute for Immunology
CH-4005 Basel, Switzerland

++Division of Biology
California Institute of Technology
Pasadena, California

§Institute for Immunology and Genetics
German Cancer Research Center
D-6900 Heidelberg, Federal Republic of Germany

Running title: Organization and evolution of the *H-2D* region class I genes

INTRODUCTION

The major histocompatibility complex (MHC) encodes highly polymorphic cell surface glycoproteins which are recognized alone, or in association with foreign antigens, on the surface of target cells by T lymphocytes (for reviews see 1-3). Most cytotoxic T cells recognize antigens in the context of MHC class I molecules, while most helper T cells do so in the context of MHC class II molecules. In the BALB/c inbred mouse strain, three class I molecules, namely K, D, and L, and two class II molecules, called A and E, have been identified. In addition to the K, D and L transplantation antigens, which are present on virtually all somatic cells, a growing number of MHC class I molecules (Qa-1, Qa-2, TL) with restricted tissue distribution have been characterized, the function of which is unknown.

Mouse strains can be distinguished from one another by their particular sets of MHC alleles or haplotypes (4). Haplotypes are indicated by lower-case letters; e.g., the BALB/c, C57BL/10, and AKR inbred strains are representatives of the d, b and k haplotypes. The MHC of the BALB/c mouse contains at least 33 class I genes, two in the *K* region, 13 in the *D* and *Qa* regions, and 13 in the *Tla* region (5-7) (Fig. 1). Among these, the K^d, D^d and L^d genes have been cloned and characterized (15,16). Two tissue-specific class I genes have been extensively characterized, namely, the *Q10* gene (17-21), coding for a liver-specific secreted molecule, and the *Tla-1,2* gene (7,12-14), coding for a T cell-specific molecule.

Serological and molecular genetic techniques have yielded conflicting results regarding the number of class I genes in the *D* region of the MHC. For instance, five D^d region class I proteins, D^d, L^d, R^d, M^d and $L2^d$, have been identified by immunochemical methods (22-24), while only three D^d region class I genes (D^d, L^d and one additional gene) have been isolated (5,6,16). It is not clear at present whether the R^d, M^d and $L2^d$ proteins are encoded by separate D^d region genes,

some of which might not have been isolated, or are derived by alternative RNA splicing or posttranslational modification from the known D^d region class I genes.

Serological and molecular evidences exist for different numbers of class I genes and gene products in the *D* and *Qa* regions of different mouse inbred strains. First, by immunoprecipitation, peptide map analysis and co-capping techniques, three to five class I molecules have been identified in the D^d and D^q, two in the D^{w16} and one each in the D^b, D^{dx}, D^k and D^p regions (22-27). Second, while three class I genes have been cloned from the D^d, only one has been isolated from the D^b region (28). Third, some mouse strains express a Qa-2 class I molecule which is encoded in the *Qa* region, while other strains do not (29). Fourth, compared to the cloned *Qa* region of the C57BL/10 mouse, two genes have apparently been lost in the BALB/c (Qa-2^+) (28) and presumably a third gene is lost in the BALB/cBy (Qa-2^-) (30).

To address the question of how many class I genes occur in the D^d region of BALB/c, a chromosomal walking technique was used to extend the previously cloned D^d region. Five class I genes are shown to be located in the D^d region, while one class I gene, identified as D^k, was isolated from the AKR mouse. Comparison of the molecular map of the D^d region to similar maps of the D^b and D^k regions allowed us to speculate on the evolutionary processes that generated the variation of numbers of *D* and *Qa* region class I genes in different mouse strains.

MATERIALS AND METHODS

Chromosome Walking

A BALB/c cosmid library, constructed using the vector pNNL (31) as previously described (32), was screened successively (see Fig. 2) with probes 13 (yielding clones II5.16 and II7.13 due to cross hybridization), 5 (yielding clone II2.20), 7

(yielding clone II6.18), 14 (yielding clone II1.10 and II4.19), 16 (yielding clones II6.17 and II3.20) and 1 + 3 (yielding clones II3.5 and II4.12). The overlap between clone II3.20 and the previously isolated clone 2.1 (5), was confirmed by hybridization with probe 17 (see Fig. 2). A BALB/c phage L library, constructed using the vector EMBL3 (33), was screened with probes 9 + 11, yielding clones L1.2 and L5.4 (see Fig. 2).

An AKR cosmid library (32) was screened successively with probes 13 (yielding clones k1.3, k2.4, k12.1, k14.2, k21.4, k13.2), 3 (yielding clone k9.4) and with probe D [a K^k region probe (32)], yielding clones k1.4, k6.2 and k11.1 due to cross hybridization (see Fig. 3).

Hybridizations were done as described (34), except that restriction fragments used as probes were isolated and labeled according to Feinberg and Vogelstein (35).

Hybridization Probes

Hybridization probes shown in Figs. 2 and 3 were derived from the following cosmid clones: 1: 2.5 kb SacII-KpnI fragment from k9.4; 2: 0.5 kb Sau3A fragment from the 4.5 kb BamHI fragment of clone dm1-38.1, described in (36); 3: 0.8 kb HpaI fragment from k1.3; 4: probe 18.1B described in (36); 5: 1.4 kb SmaI fragment from II7.13; 6: 0.5 kb AluI fragment from the 3.5 kb HpaI fragment of II2.20; 7: 6.2 kb HpaI fragment from II2.20; 8: 0.2-0.5 kb HpaI-HindIII fragment from the 1 kb SmaI-SalI fragment of II6.18 (the SalI site is located in the cosmid vector); 9: 1.8 kb SmaI fragment from 50.2; 10: 0.9 kb Sau3A fragment from the 9 kb XhoI-NruI fragment of BALB/c cosmid clone JS4.4 (H.S., unpublished; the NruI site is located in the cosmid vector) extending from map position 185 to 226 kb (see Fig. 2); 11: DpnI fragment from the 5.2 kb NruI-SmaI fragment of 16.1 (the NruI site is located in the cosmid vector); 12: probe 59.2B described in (36); 13: probe 27.51 described in (5); 14: 1.2 kb SmaI fragment from 59.2; 15: 1 kb KpnI-SalI fragment

from II1.10 (the SalI site is located in the cosmid vector); 16: 3.2 kb SmaI-SalI fragment from II4.19 (the SalI site is located in the cosmid vector); 17: 1.1 kb BamHI fragment from 2.1; 18: DpnI fragment from the right end of the k11.1 (see Fig. 3).

D^k Gene Transfer and Expression

Transfections were performed as described previously (37). For some transfections, the plasmid clone pDK17, constructed by subcloning the 11 kb EcoRI fragment of cosmid K12.1 into pBR322, was used. For transfection, the (C3H x C57BL/6)F_1 mouse fibrosarcoma cell line IC9 that expresses only the D^b molecule (37) or a B10.CAS2 mouse fibroblast cell line was used. Briefly, for the IC9 cell line class I DNA (10-20 µg) and pAG60 (1-4 µg) (38) and for the B10.CAS2 cell line class I DNA and pSV2neo (39) was coprecipitated with calcium phosphate and added to about 10^6 adherent cells. After 5 hr at 37°C the medium was replaced by 5 ml 15% glycerol for 2 min to facilitate uptake of DNA by the cells. Following overnight incubation with medium, the transfected IC9 cells were collected with EDTA buffer and distributed into 2 Costar 3524 multiwell plates in selection medium containing 700 µg/ml G418 (Gibco). B10.CAS2 transfectants were selected with 1 mg/ml G418 and tested uncloned. IC9 transfectants were expanded, tested for expression of the D^k molecule and then subcloned. H-2 expression on IC9 transfectants was determined in a cellular radioimmunoassay using ^{125}I-labeled protein A, as described previously (40). Briefly, 10^6 cells were incubated in microtiter plates with specific anti-H-2 monoclonal antibody. After washing, ^{125}I-labeled protein A was added and the amount of bound radioactivity determined. Cpm values represent the mean of triplicate wells. The standard deviations range between 5% and 10% (not shown). The following monoclonal anti-H-2 antibodies were used: H142-23, H100-27, H100-30, B22-249, H116-22, 18/20 (40); K7-309 (41); R1-21.2 (42); 15-5-5s, 16-1-2 (43). H11-3 was a generous gift of Dr. Svendsen, Copenhagen (Statens Seruminstitut).

RESULTS

500 kb of Cloned BALB/c DNA Links the *D* and *Qa* Regions at the Molecular Level and Contains Five *D* Region and Eight *Qa* Region Class I Genes

We previously described five BALB/c class I gene clusters, termed clusters 1, 2, 6, 9 and 13, that were mapped into the *D* and *Qa* regions of the MHC and together contained 13 class I genes (5,6). The single class I gene on cluster 13 and one of the two class I genes on cluster 2 encoded the D^d and L^d proteins, respectively (16). Further characterization showed that clusters 1 and 9, both mapping into the *Qa* region, are overlapping (44). To link the *D* and *Qa* region gene clusters at the molecular level, we have now isolated a series of overlapping phage lambda and cosmid clones from BALB/c liver DNA libraries by using hybridization probes derived from previously isolated BALB/c clones and from the 5' flanking sequence of the D^k gene (see below and Materials and Methods). Three chromosome walking steps were necessary to link clusters 13 and 6, one to link cluster 6 and 2, and two to link clusters 2 and 1+9. Screening of the BALB/c cosmid library with probes from the 5' flanking region of the L^d and D^k genes led to the isolation of about 80 kb of DNA upstream of the D^d gene (Fig. 2).

The previously described clones together with the newly isolated ones allowed us to define a stretch of 500 kb of DNA which spans the *D* and *Qa* regions of the MHC of the BALB/c mouse (Fig. 2). This stretch contains 13 class I genes, identified with the highly crossreactive class I cDNA clone pH-2IIa (5), namely, five in the *D* region and eight in the *Qa* region. The three class I genes located between D^d and L^d are provisionally called $D2^d$, $D3^d$ and $D4^d$. The eight class I genes in the *Qa* region are denoted *Q1*, *Q2*, *Q4-Q7*, *Q8/9* and *Q10*, in agreement with the nomenclature established by Weiss *et al.* (28) for the *Qa* region class I genes in the C57BL/10 mouse. It appears that the *Q3* gene has been lost from the BALB/c genome and that the *Q8* and *Q9* genes have fused into a *Q8/9* hybrid gene

(28). Gene *Q7* corresponds to the previously sequenced gene 27.1 (5,45).

The stretch of 500 kb of DNA is defined by a total of 84 overlapping cosmid clones, 28 of which are shown in Fig. 2, and six phage lambda clones, two of which are shown. Individual cosmid clones overlap by more than 10 kb, except for a gap between cosmid clones 50.2 and II 4.8, which is bridged by the lambda clones 1.2 and 5.4. These clones were isolated with probes 9 and 11 and share restriction maps with the cosmid clones as shown. The linkage was further confirmed by mapping BamHI, XbaI, SstI and HindIII sites in the two lambda clones and the overlapping cosmid clones (not shown).

130 kb of Cloned DNA Containing Two Class I Genes Link the *D* and *Qa* Regions of the AKR Mouse at the Molecular Level

To compare the *D* region gene organization in BALB/c and AKR mouse strains, we screened an AKR cosmid library with probe 13, isolated from the 5' flanking region of the L^d gene (see Materials and Methods). Six overlapping cosmid clones were obtained which defined a stretch of 61 kb of DNA with a single class I gene (Fig. 3). The cloned region was extended to the 5' side of the class I gene by a chromosome walk using probe 3, leading to the isolation of clone k9.4 (Fig. 3). The 3' side was extended with three cosmid clones, k1.4, k6.2 and k11.1 (Fig. 3), which had been isolated from the AKR cosmid library with probe D from the 3' flanking region of the K^k gene due to crosshybridization (see 22). Clone k11.1 contained a second class I gene as revealed by hybridization to the cDNA clones pH-2III and pH-2IIa which define the 5' and 3' half of class I genes, respectively (5) (Fig. 3). Comparison of the restriction maps of the two AKR class I genes with those of the BALB/c *D* and *Qa* region class I genes showed that the upstream AKR gene is homologous to D^d and L^d, while the downstream AKR gene is homologous to the *Q1* gene of BALB/c (compare Figs. 2 and 3). This result suggested that the upstream AKR gene is D^k and the downstream gene *Q1*. To confirm that the

cloned region does indeed link the AKR *D* and *Qa* regions at the molecular level, we mapped the two single-copy probes 1 and 18 with respect to their location on the genetic map of the MHC. As shown in Fig. 4, the restriction fragment length polymorphism identified with probe 1 maps to the *D* region, while the polymorphic restriction fragment identified with probe 18 maps to the *Qa* region. Thus, we conclude that the cloned stretch of 130 kb of DNA covers part of the *D* and *Qa* region of the AKR mouse.

Identification of the D^k Gene by Gene Transfer and Expression

To confirm that the proximal class I gene in the 130 kb gene cluster corresponds to the D^k gene, we cotransfected the neomycin resistance gene together with each of the cosmid clones k1.3 and k12.1 into a B10.CAS2 mouse fibroblast cell line and the same neomycin resistance gene together with cosmid clone k13.2 as well as a plasmid subclone (pDk17) of k12.1 into the mouse fibrosarcoma cell line IC9 which expresses only the D^b molecule (see Materials and Methods). Transfectants were selected in the presence of G418 and tested, either uncloned or cloned, by radioimmunoassays with a panel of monoclonal antibodies and with D^k-alloreactive cytotoxic T cells (Tables 1 and 2). Clearly, three of the cosmid clones, when introduced into the cell lines, led to the appearance of D^k serological determinants on the surface of the transfected cells. Transfected IC9 cells were positive for all monoclonal antibodies with reactivity against the D^k molecule. Furthermore, B10.CAS2 fibroblasts, when transfected with cosmid clones k1.3 or k12.1, were killed by the D^k-specific cytotoxic T cells. Together with the genetic and physical mapping data we can conclude that we have correctly identified the D^k gene.

Four of the Five *D* Region Genes of BALB/c Mice Appear To Be Missing in Some Strains and Can Undergo Deletions

A comparison of the *D* region gene organization in the three mouse strains which

have been characterized by molecular cloning, BALB/c, AKR and C57BL/10, is shown in Fig. 5. While five class I genes have been identified in the *D* region of BALB/c, only one has been found in AKR and in C57BL/10. The D^d, D^k, D^b and L^d genes have very similar restriction maps for their flanking sequences (Fig. 5), although the homology region 5' of the L^d gene is short.

The cloning studies indicated that four of the five BALB/c *D* region class I genes are missing in AKR and C57BL/10. Southern blot analysis confirmed that sequences located between D^d and L^d genes are absent from AKR and C57BL/10 DNA. As shown in Fig. 6, hybridization probes 7 and 11, isolated from the region between D^d and L^d, do not detect complementary sequences in AKR and C57BL/10 DNA. On the other hand, probes 3 and 15, isolated from the 5' flanking region of D^d and 3' region of L^d, respectively, detect the same restriction fragments in AKR and C57BL/10 as in BALB/c DNA. The 4.8 kb BamHI fragment detected by probe 7 in all three mouse DNA is due to crosshybridization of this probe to a sequence, labeled "7 cross," located 3' to the L^d gene (Fig. 6 and see below). This Southern blot analysis indicates that the D^p and D^q regions in mouse strains B10.F(14R) and B10.AKM, respectively, with the four *D* region hybridization probes show that the D^q region is indistinguishable from the D^d region, while in the D^p region sequences 11 and "7 cross" are apparently deleted (Fig. 6).

With the *D* region probes we have also analyzed three d haplotype mutant strains, one with a mutation in the *K* locus (C57BL/6-H-2^{dm5}) (50) and two with mutations in the *D* region (B10.D2-H-2^{dm1} and BALB/c-H-2^{dm2}). The dm1 mutant has recently been shown to contain a D^d/L^d hybrid gene (36,51) while the dm2 mutant does not express the L^d antigen (52). As shown in Fig. 6, the *K* region mutant dm5 does not show any gross changes from BALB/c, but the two *D* region mutants show deletion of probes 7 and 11. This is consistent with the previous finding that the dm1 mutant has deleted the region between D^d and L^d, presum-

ably due to unequal crossing-over between the D^d and L^d genes, thus forming the hybrid gene (36). The present data indicate that the dm2 mutation might also be due to unequal crossing-over, deleting at least $D2^d$ and $D3^d$. Since the mutant fails to express L^d, it is likely that the deletion extends into the L^d gene.

DNA Sequences from the D^d Region Cross Hybridize Most Strongly to Other D^d and to *Qa* Region Sequences

We used 11 restriction fragments isolated from the *D* region as hybridization probes to analyze the relationship of the D^d region DNA sequences to those of the other class I regions of the MHC of the BALB/c mouse. The 11 restriction fragments were labeled (35) and then hybridized to dot blots containing 33 cosmid clones which covered a total of 1000 kb and represented all cloned class I regions (not shown). Cosmid clones giving a positive signal were then analyzed by Southern blotting with the appropriate probes. Although the results obtained were complex, certain recurrent patterns can be seen (Fig. 7). First, the four probes, 4, 6, 7, and 8, isolated from the segment between genes D^d and $D2^d$ all cross hybridize to sequences located between genes L^d and *Q1* and the two probes, 14 and 15, isolated from the region between the L^d and *Q1* genes cross hybridize to the D^d-$D2^d$ segment. The D^d-$D2^d$/L^d-*Q1* homologies as seen by cross hybridization are in perfect agreement with the restriction map homologies mentioned above. Second, the two probes, 9 and 10, isolated from the $D3^d$-$D4^d$ segment cross hybridize to the *Qa* region segment *Q1-Q4*, a relationship which was not seen by restriction map comparison. Third, the D^d-$D2^d$ segment probe, 8, also shows strong hybridization to the *Q5-Q10* genes as well as to the K^d and $K2^d$ genes. Fourth, the $D4^d$-L^d segment probe, 12, cross hybridizes about 21 kb upstream of the D^d gene, while the two probes, 2 and 3, located 32-40 kb upstream of the D^d gene shown no cross hybridization to the 5' flanking sequence of the L^d gene and, in fact, are single-copy probes. These sequence homologies have important evolutionary implications as will be discussed below.

DISCUSSION

Organization and Expression of Class I Genes in the *D* and *Qa* Regions of the BALB/c Mouse

Thirteen class I genes have been linked into a contiguous stretch of 500 kb of DNA extending from the *D* region to the *Qa* region of the MHC of the BALB/c mouse. We have previously shown that low copy-number probes isolated from the 5' flanking region of the D^d and from the 3' flanking region of the L^d gene, map into the *D* region; a probe isolated from the 3' flanking region of the *Q2* gene maps into the Qa region; and a probe from the 3' flanking region of the *D3* gene maps to either *D* or *Qa* region (6). These results are in agreement with the molecular linkage map. The localization of genes $D2^d$ and $D3^d$ (cluster 6) between D^d and L^d is also in agreement with their deletion in the dm1 and dm2 mutants. Weiss *et al.* (28) have mapped the recombination point in the B6.K1 recombinant strain that separates the *D* from the *Qa* region to within a 45 kb region between D^b and *Q1*. Thus, there are at least five class I genes in the D region and eight in the *Qa* region of the BALB/c mouse. The localization of the D^d gene proximal to L^d by molecular linkage is consistent with the gene order recently determined by genetic recombination (53).

Two of the five D^d region class I genes code for the D^d and L^d antigens(46-48). Whether the $D2^d$, $D3^d$ and $D4^d$ genes are functional genes remains to be determined. The identification of these three additional D^d region class I genes should readily permit their characterization to find out whether they encode any of the additional D^d region class I proteins, namely, M^d, R^d and $L2^d$ (22-24). Five *D* region class I proteins have also been identified (25) in mouse strains carrying the D^q region which, according to Southern blot analysis (Fig. 6), appears to have a similar gene organization as the D^d region. The fact that these additional genes are missing in some mouse strains and that they can be deleted in the dm1 and

dm2 mutants without apparent deleterious phenotypic effect indicates that they are not essential, at least under laboratory conditions.

Of the eight class I genes located in the Qa region of the BALB/c mouse, four, *Q6*, *Q7*, *Q8/9* and *Q10*, have been carefully analyzed for their expression in mouse L cells after gene transfer (54,55). None of them could be shown to encode a cell surface class I molecule (54,55), in contrast to a previous study (16). The lack of cell surface expression of these four genes appears to be due to information encoded by the 3' portion of these genes (54,55). Only *Q8/9*, however, appears to be a truly nonfunctional gene, while *Q6*, *Q7* and *Q10* might encode secreted class I proteins or are under tissue-specific regulation which prevented their stable cell surface expression in L cells (54,55). Indeed, the *Q10* gene of C57BL/10 mouse has been shown to encode a liver-specific, secreted class I protein (17-19). Moreover, all of the *Q6*, *Q7*, *Q8*, and *Q9* genes of the C57BL/10 mouse appear to encode class I proteins containing Qa-2 serological determinants (30).

Duplication of *D* Region Class I Genes By Unequal Crossing-over

Three recombinational mechanisms, in addition to the accumulation of point mutations, contribute to the generation and maintenance of polymorphic DNA sequences in the class I genes. First, gene conversion appears to be the primary genetic mechanism generating the variability in the K^b gene and, presumably, other class I genes as well (56-59). Second, homologous equal crossing-over events, occurring with high preference at certain sites termed recombinational hot spots, lead to a relatively rapid shuffling of chromosomal segments of the MHC (8,60). Recombination at hot spots therefore contributes to the generation of new MHC haplotypes and might be important for the maintenance of useful MHC alleles during evolution. Third, homologous unequal crossing-over events apparently generate class I gene duplications and deletions as seen in the *Qa* region (28,30) and in the dm1 mutant (36).

The homology, as seen with restriction maps, between the flanking regions of the D^d, L^d, D^b and D^k genes, the homologies detected by cross hybridization of probes between the flanking sequences of genes $D2^d$, $D3^d$, $D4^d$ and the flanking sequences of the *Qa* region genes, together with the observations that the spacings between D^d, $D2^d$, $D3^d$ and $D4^d$ are similar to those between L^d, *Q1*, *Q2* and *Q4*, suggested that the first four D^d region genes were duplicated through unequal crossing-over. However, while the D^d gene is highly homologous to the L^d, the *D2-D4* region is much less homologous to its counterpart, the *Q1-Q4* region. The evolution of these genes probably happened in two steps as shown in Fig. 7. The ancestral *D* gene first exchanged by homologous recombination with its allelic locus on the homologous chromosome, which came from a distantly related parent and was much diverged. The resulting recombinant chromosome was then recombined, through unequal crossing-over, with the original BALB/c-like haplotype chromosome, generating an expanded chromosome with five class I genes in the *D* region, three of which were derived from the *Qa* region (Fig. 8).

If the proposed evolutionary origin of the BALB/c *D* region genes is correct, then the presence of a single class I gene in mouse strains AKR and C57BL/10 reflects the, evolutionarily speaking, older *D* region gene organization. However, it is, of course, also conceivable that some of the present day strains, which contain a single *D* region class I gene, originated through gene deletions from strains with multiple D region class I genes. The dm1 and dm2 mutants clearly are examples of this process.

It is reasonable to assume that the many examples of unequal crossing-over in the *D* and *Qa* regions are due to the close spacing, the sequence homologies and the identical *5'* to *3'* orientation of the multiple class I genes in this part of the murine MHC.

SUMMARY

Chromosome walking has been used to link the *D* and the *Qa* regions of the MHC of the BALB/c mouse and also to clone part of the *D* region of the AKR mouse. At least five class I genes are located in the *D* region of the BALB/c mouse, while only a single class I gene, which has been identified by transfection as the D^k gene, has been found in the *D* region of the AKR mouse. Comparison of the molecular map of the D^d region to maps of the D^b and D^k regions and cross hybridizations of flanking probes suggest that the multiple class I genes in the D^d region have been generated by unequal crossing-over involving class I genes from the *Qa* region. The expanded D^d region appears to be metastable, since evidence for gene family contraction in the D^d region has been found in two mutant strains. Thus, the *D* region are in a dynamic state, evolving by gene family expansion and contraction.

REFERENCES

1. Klein, J., F. Figueroa, and Z.A. Nagy. 1983. Genetics of the major histocompatibility complex: The final act. Ann. Rev. Immunol. **1**, 119-142.
2. Schwartz, R.H. 1985. T-lymphocyte recognition of antigen in association with gene products of the major histocompatibility complex. Ann. Rev. Immunol. **3**, 237-261.
3. Steinmetz, M., and L. Hood. 1983. Genes of the major histocompatibility complex in the mouse and man. Science **222**, 727-733.
4. Klein, J., F. Figueroa, and C.S. David. 1983. *H-2* haplotypes, genes, and antigens: Second listing. II. The *H-2* complex. Immunogenetics **17**, 553-596.
5. Steinmetz, M., A. Winoto, K. Minard, and L. Hood. 1982. Clusters of genes encoding mouse transplantation antigens. Cell **28**, 489-498.
6. Winoto, A., M. Steinmetz, and L. Hood. 1983. Genetic mapping in the major histocompatibility complex by restriction enzyme site polymorphisms: Most mouse class I genes map to the *Tla* complex. Proc. Natl. Acad. Sci. USA **80**, 3425-3429.
7. Fisher, D.A., S.W. Hunt III, and L.E. Hood. 1985. Structure of a gene encoding a murine thymus leukemia antigen and organization of *Tla* genes in the BALB/c mouse. J. Exp. Med. **162**, 528-545.
8. Steinmetz, M.. D. Stephan, and K. Fischer Lindahl. 1986. Gene organization and recombinational hot spots in the murine major histocompatibility complex. Cell, in press.
9. Chaplin, D.D., D.E. Woods, A.S. Whitehead, G. Goldberger, H.R. Colten, and J.G. Seidman. 1983. Molecular map of the murine *S* region. Proc. Natl. Acad. Sci USA **80**, 6947-6951.

10. Amor, M., M. Tosi, C. Duponchel, M. Steinmetz, and T. Meo. 1985. Liver mRNA probes disclose two cytochrome P-450 genes duplicated in tandem with the complement C4 loci of the mouse *H-2 S* region. Proc. Natl. Acad. Sci. USA **82**, 4453-4457.

11. White, P.C., D.D. Chaplin, J.H. Weis, B. Dupont, M.I. New, and J.G. Seidman. 1984. Two steroid 21-hydroxylase genes are located in the murine *S* region. Nature **312**, 465-467.

12. Obata, Y., Y.-T. Chen, E. Stockert, and L.J. Old. 1985. Structural analysis of TL genes of the mouse. Proc. Natl. Acad. Sci. USA **82**, 5475-5479.

13. Chorney, M.J., J.S. Tung, Y. Bushkin, and F.W. Shen. 1985. Structural characteristics of *Tla* products. J. Exp. Med. **162**, 781-789.

14. Chen, Y.-T., Y. Obata, E. Stockert, and L.J. Old. 1985. Thymus-leukemia (TL) antigens of the mouse. Analysis of TL mRNA and TL cDNA from TL+ and TL- strains. J. Exp. Med. **162**, 1134-1148.

15. Goodenow, R.S., M. McMillan, A. Orn, M. Nicolson, N. Davidson, J.A. Frelinger, and L. Hood. 1982. Identification of a BALB/c $H\text{-}2L^d$ gene by DNA-mediated gene transfer. Science **215**, 677-679.

16. Goodenow, R.S., M. McMillan, M. Nicolson, B.T. Sher, K. Eakle, N. Davidson, and L. Hood. 1982. Identification of the class I genes of the mouse major histocompatibility complex by DNA-mediated gene transfer. Nature **300**, 231-237.

17. Cosman, D., M. Kress, G. Khoury, and G. Jay. 1982. Tissue-specific expression of an unusual *H-2* (class I)-related gene. Proc. Natl. Acad. Sci, USA **79**, 4947-4951.

18. Maloy, W.L., J.E. Coligan, Y. Barra, and G. Jay. 1984. Detection of a secreted form of the murine *H-2* class I antigen with an antibody against its predicted carboxyl terminus. Proc. Natl. Acad. Sci. USA **81**, 1216-1220.

19. Mellor, A.L., E.H. Weiss, M. Kress, G. Jay, and R.A. Flavell. 1984. A nonpolymorphic class I gene in the murine major histocompatibility complex. Cell **36**, 139-144.

20. Devlin, J.J., A.M. Lew, R.A. Flavell, and J.E. Coligan. 1985. Secretion of a soluble class I molecule encoded by the *Q10* gene of the C57BL/10 mouse. EMBO J. **4**, 369-374.

21. Barra, Y., K. Tanaka, K.J. Isselbacher, G. Khoury, and G. Jay. 1985. Stable transfer and restricted expression of a cloned class I gene encoding a secreted transplantation-like antigen. Mol. Cell. Biol. **5**, 1295-1300.

22. Hansen, T.H., K. Ozato, M.R. Melino, J.E. Coligan, T.J. Kindt, J. Jandinski, and D.H. Sachs. 1981. Immunochemical evidence in two haplotypes for at least three *D* region-encoded molecules, D, L, and R. J. Immunol. **126**, 1713-1716.

23. Ivanyi, D., and P. Demant. 1979. Complex genetic effect of B10.D2(M504)-(H-2^{dm1}) mutation. Immunogenetics **8**, 539-550.

24. Ivanyi, D., and P. Demant. 1982. Molecular heterogeneity of *D*-end products detected by anti-H2.28 sera. II. B10.D2 (M504)(H-2^{dml}) mutant fails to express one of the two H-2.4^-, H-2.28^+ D^d region molecules. Immunogenetics **15**, 467-476.

25. Ivanyi, D., and P. Demant. 1984. Five serologically distinguishable D^q region molecules. Immunogenetics **20**, 211-216.

26. Chatterjee-Das, S., S.K. Berg, and D.H. Sachs. 1985. Molecular and serologic analysis of products of the *D* region of H-2^q. Transplant. Proc. **17**, 722-724.

27. Lillehoj, E.P., N.B. Myers, D.R. Lee, T.H. Hansen, and J.E. Coligan. 1985. Structural definition of a family of L^d-like molecules distributed among four of seven haplotypes compared. J. Immunol. **135**, 1271-1275.

28. Weiss, E.H., L. Golden, K. Fahrner, A.L. Mellor, J.J. Devlin, H. Bullman, H. Tiddens, H. Bud, and R.A. Flavell. 1984. Organization of the class I gene family in the major histocompatibility complex of the C57BL/10 mouse. Nature **310**, 650-655.

29. Flaherty, L. 1981. *Tla*-region antigens. In: ***Role of the Major Histocompatibility Complex in Immunology***, ed. M. Dorf (Garland, New York), pp. 33-57.

30. Mellor, A.L., D. Antoniou, and P.J. Robinson. 1985. Structure and expression of genes encoding murine Qa-2 class I antigens. Proc. Natl. Acad. Sci. USA **82**, 5920-5924.

31. Grosveld, F.G., T. Lund, E.J. Murray, A.L. Mellor, H.H.M. Dahl, and R.A. Flavell. 1982. The construction of cosmid libraries which can be used to transform eukaryotic cells. Nucleic Acids Res. **10**, 6715-6732.

32. Steinmetz, M., M. Malissen, L. Hood, A. Orn, R.A. Maki, G.R. Dastoornikoo, R. Stephan, E. Gibb, and R. Romaniuk. 1984. Tracts of high or low sequence divergence in the mouse major histocompatibility complex. EMBO J. **3**, 2995-3003.

33. Frischauf A.-M., H. Lehrach, A. Poustka, and N. Murray. 1983. Lambda replacement vectors carrying polylinker sequences. J. Mol. Biol. **170**, 827-842.

34. Steinmetz, M., D. Stephan, G.R. Dastoornikoo, E. Gibb, and R. Romaniuk. 1985. Methods in molecular immunology: Chromosomal walking in the major histocompatibility complex. In: ***Immunological Methods***, Vol. III, eds. I. Lefkovits, and B. Pernis (Academic Press, New York), pp. 1-19.

35. Feinberg, A.P., and B. Vogelstein. 1984. Addendum: A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **137**, 266-267.

36. Sun, H., R.S. Goodenow, and L. Hood. 1985. Molecular basis of the dml mutation in the major histocompatibility complex of the mouse: A *D/L* hybrid gene. J. Exp. Med. **162**, 1588-1602.

37. Wallich, R., N. Bulbuc, G.J. Hammerling, S. Katzav, S. Segal, and M. Feldman. 1985. Abrogation of metastatic properties of tumor cells by *de novo* expression of H-2K antigens following *H-2* gene tranfection. Nature **315**, 301-305.

38. Colbere-Garapin, F., F. Horodniceanu, P. Kourilsky, and A.-C. Garapin. 1980. A new dominant hybrid selective marker for higher eukaryotic cells. J. Mol. Biol. **150**, 1-14.

39. Southern, P.J. and P. Berg. 1982. Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. J. Mol. Appl. Genet. **1**, 327-341.

40. Lemke, H., G.J. Hammerling, and U. Hammmerling. 1979. Fine specificity analysis with monoclonal antibodies of antigens controlled by the major histocompatibility complex and by the Qa/TL region in mice. Immunol. Rev. **47**, 175-206.

41. Hammmerling, G.J., E. Rusch, N. Tada, S. Kimura, and U. Hammerling. 1982. Localization of allodeterminants on H-2K^{b} antigens determined with monoclonal antibodies and H-2 mutant mice. Proc. Natl. Acad. Sci. USA **79**, 4737-4741.

42. Koch, S., N. Koch, P. Robinson, and G.J. Hammerling. 1983. Comparison of allo- and xenogeneic determinants on the H-2K^{k} molecule. Transplantation **36**, 177-180.

43. Ozato, K., N. Mayer, and D. Sachs. 1980. Hybridoma cell line secreting monoclonal antibodies to mouse H-2 and Ia antigens. J. Immunol. **124**, 533-540.

44. Rogers., J.H. 1985. Family organization of mouse *H-2* class I genes. Immunogenetics **21**, 343-353.

45. Steinmetz, M., K.W. Moore, J.G. Frelinger, B.T. Sher, F.-W. Shen, E.A. Boyse, and L. Hood. 1981. A pseudogene homologous to mouse transplantation antigens: transplantation antigens are encoded by eight exons that correlate with protein domains. Cell **25**, 683-692.

46. Sher, B.T., R. Nairn, J.E. Coligan, and L. Hood. 1984. DNA sequence of the mouse $H\text{-}2D^d$ transplantation gene. Proc. Natl. Acad. Sci. USA **82**, 1175-1179.

47. Moore, K.W., B.T. Sher, Y.H. Sun, K.A., Eakle, and L. Hood. 1982. DNA sequence of a gene encoding a BALB/C mouse L^d transplantation antigen. Science **215**, 679-682.

48. Evans, G.A., D.H. Margulies, R.D. Camerini-Otero, K. Ozato, and J.G. Seidman. 1982. Structure and expression of a mouse major histocompatibility antigen gene, $H\text{-}2L^d$. Proc. Natl. Acad. Sci. USA **79**, 1994-1998.

49. Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

50. Melvold, R.W., M.S. Sandrin, and I,F.C. McKenzie. 1982. The C. B6-H-2^{dm5} mutant. Immunogenetics **15**, 331-332.

51. Burnside, S.S., P. Hunt, K. Ozato, and D.W. Sears. 1984. A molecular hybrid of the $H\text{-}2D^d$ and $H\text{-}2L^d$ genes expressed in the dml mutant. Proc. Natl. Acad. Sci. USA **81**, 5204-5208.

52. Hansen, T.H., S.E. Cullen, R. Melvold, H.I. Kohn, L. Flaherty, and D.H. Sachs. 1977. Mutation in a new H-2-associated histocompatibility gene closely linked to *H-2D*. J. Exp. Med. **145**, 1550-1558.

53. Savarirayan, S., W.P. Lafuse, and C.S. David. 1985. Recombination between *H-2D* and *H-2L* genes: Identification, characterization, and gene order. Transplant. Proc. **17**, 702-706.

54. Stroynowski, I., J. Forman, R.S. Goodenow., S.G. Schiffer, M. McMillan, S.O. Sharrow, D.H. Sachs, and L. Hood. 1985. Expression and T cell recognition of hybrid antigens with amino-terminal domains encoded by *Qa-2* region of major histocompatibility complex and carboxyl termini of transplantation antigens. J. Exp. Med. **161**, 935-952.

55. Straus, D.S., I. Stroynowski, S.G. Schiffer, and L. Hood. 1985. Expression of hybrid class I genes of the major histocompatibility complex in mouse L cells. Proc. Natl. Acad. Sci. USA **82**, 6245-6249.

56. Weiss, E.H., A. Mellor, L. Golden, K. Fahrner, E. Simpson, J. Hurst, and R.A. Flavell. 1983. The structure of a mutant *H-2* gene suggests that the generation of polymorphism in *H-2* genes may occur by gene conversion-like events. Nature **301**, 671-674.

57. Mellor, A.L., E.H. Weiss, K. Ramachandran, and R.A. Flavell. 1983. A potential donor gene for the bm1 gene conversion event in the C57BL mouse. Nature **306**, 782-795.

58. Pease, L.R., D.H. Schulze, G.M. Pfaffenback, and S.G. Nathenson. 1983. Spontaneous H-2 mutants provide evidence that a copy mechanism analogous to gene conversion generates polymorphism in the major histocompatibility complex. Proc. Natl. Acad. Sci. USA **80**, 242-246.

59. Schulze, D.H., L.R. Pease, S.S. Geier, A.A. Reyes, L.A. Sarmiento, R.B. Wallace, and S.G. Nathenson. 1983. Comparison of the cloned H-2Kbm1 variant gene with the *H-2K*^{b} gene shows a cluster of seven nucleotide differences. Proc. Natl. Acad. Sci. USA **80**, 2007-2011.

60. Steinmetz, M., K. Minard, S. Horvath, J. McNicholas, J. Frelinger, C. Wake, E. Long, B. Mach, and L. Hood. 1982. A molecular map of the immune response region from the major histocompatibility complex of the mouse. Nature **300**, 35-42.

61. Reyes, A.A., M. Schold, and R.B. Wallace. 1982. The complete amino acid sequence of the murine transplantation antigen H-2D^b as deduced by molecular cloning. Immunogenetics **16**, 1-9.

Figure 1. Molecular map of the MHC of the BALB/c mouse. The six gene clusters, containing a total of 46 genes and encompassing about 1600kb of DNA, are correlated with the genetic regions and subregions of the MHC defined by recombination events. A 600kb gene cluster spans the *K* and the *I* region and contains two class I (*K2, K*) and seven class II ($A_{\beta3}$, $A_{\beta2}$, A_{β}, A_{α}, E_{β}, $E_{\beta2}$, E_{α}) genes (8). Following a gap of unknown size a 250 kb cluster contains four complement-related (*C2, Bf, Slp, C4*) and two 21-hydroxylase (*21-OH*) genes (9-11 and Steinmetz, unpublished results). This is followed at an unknown distance by the 500 kb class I gene cluster described in this paper. Finally three gene clusters with 18 class I genes have been cloned from the *Tla* region (5-7) and four *Tla* genes, including the *Tla-1,2* gene encoding the TL 1^{+} 2^{+} 3^{-} 4^{-} molecule, have been identified (7,12-14).

THE MHC OF THE BALB/c MOUSE

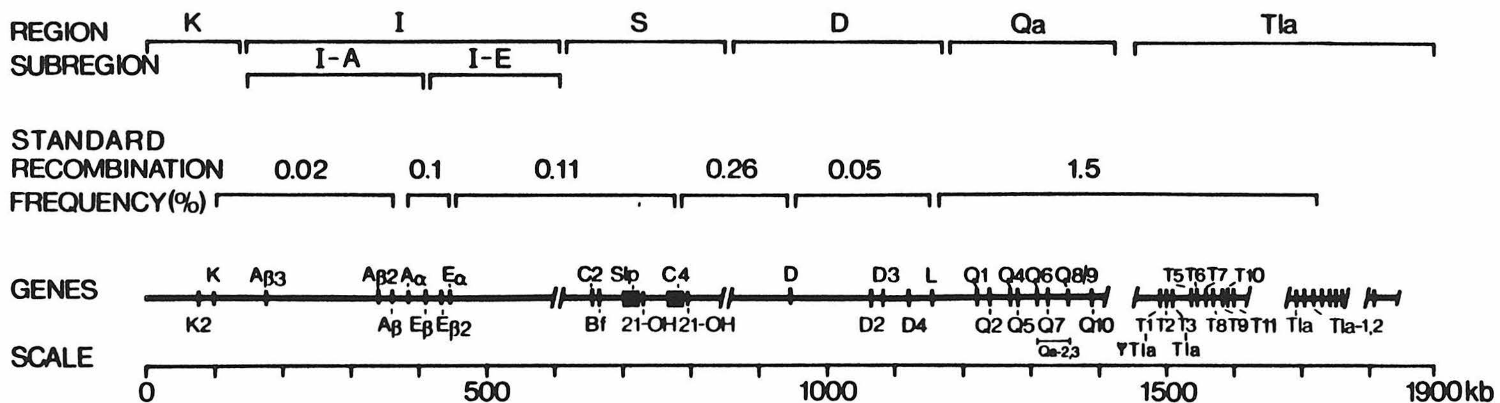
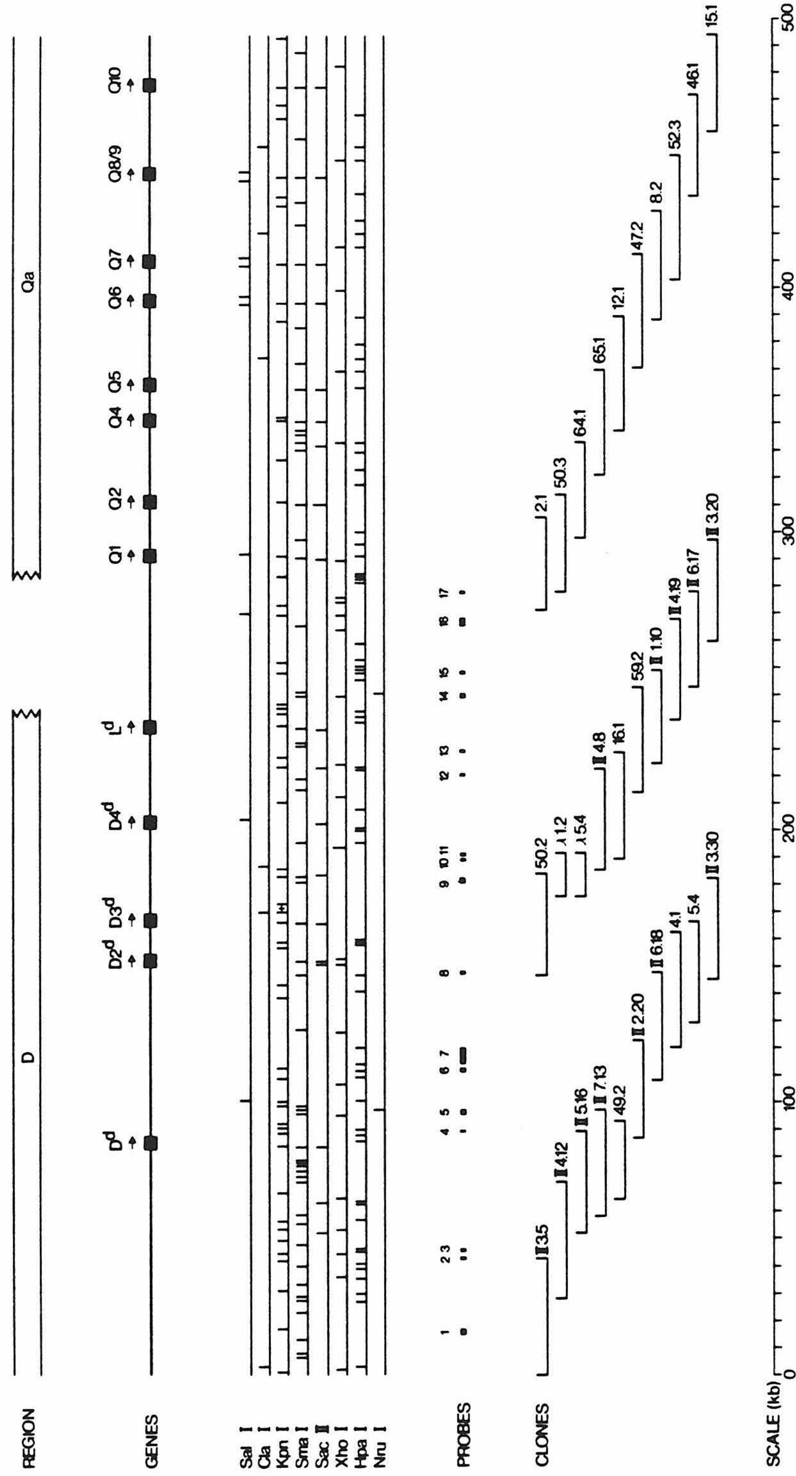


Figure 2. A gene cluster of 500 kb with 13 class I genes links the D^d and *Qa* regions of the BALB/c mouse. Genes are shown as boxes, but are not drawn to scale. The D^d, L^d, *Q7* and *Q10* genes have been sequenced (19,45-48). Restriction maps were determined by single and double (49) and by partial digestions (see 36). A "+" between two sites indicates the presence of more cleavage sites for the same enzyme. The hybridization probes were isolated as described in Materials and Methods. Cosmid and phage clones II3.35 to II7.13 (from left to right), II2.20, II6.18, II3.30, 1.2, 5.4, II4.8 and II1.10 to II3.20 were isolated in the course of this work, while cosmid clones 49.2 containing the D^d gene (36,46), cosmid clones 4.1, 5.4, 50.2 constituting cluster 6 (5), cosmid clones 16.1, 59.2 constituting cluster 2 (5), cosmid clones 2.1 to 46.1 constituting cluster 1 (5), and cosmid clone 15.1 constituting cluster 9 (5) have been described previously. The linkage between clusters 1 and 9 was described previously (44).



REGION
D
Qa
GENES
D^d
D2^d
D3^d
D4^d
L^d
Q1
Q2
Q4
Q5
Q6
Q7
Q8/9
Q10
Sal I
Cla I
Kpn I
Sma I
Sac II
Xho I
Hpa I
Nru I
PROBES
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
CLONES
II 3.5
II 4.12
II 5.16
II 7.13
49.2
II 2.20
II 6.18
4.1
5.4
II 3.30
50.2
λ 1.2
λ 5.4
II 4.8
16.1
59.2
II 1.10
II 4.19
II 6.17
II 3.20
2.1
50.3
64.1
65.1
12.1
47.2
8.2
52.3
46.1
15.1
SCALE (kb)
0
100
200
300
400
500

Figure 3. A 130 kb gene cluster with two class I genes links the D^k and *Qa* regions of the AKR mouse. Genes, restriction maps and hybridization probes are shown as in Fig. 2 Cosmid clones k9.4 to k11.1 (from left to right) were isolated as described in Materials and Methods.

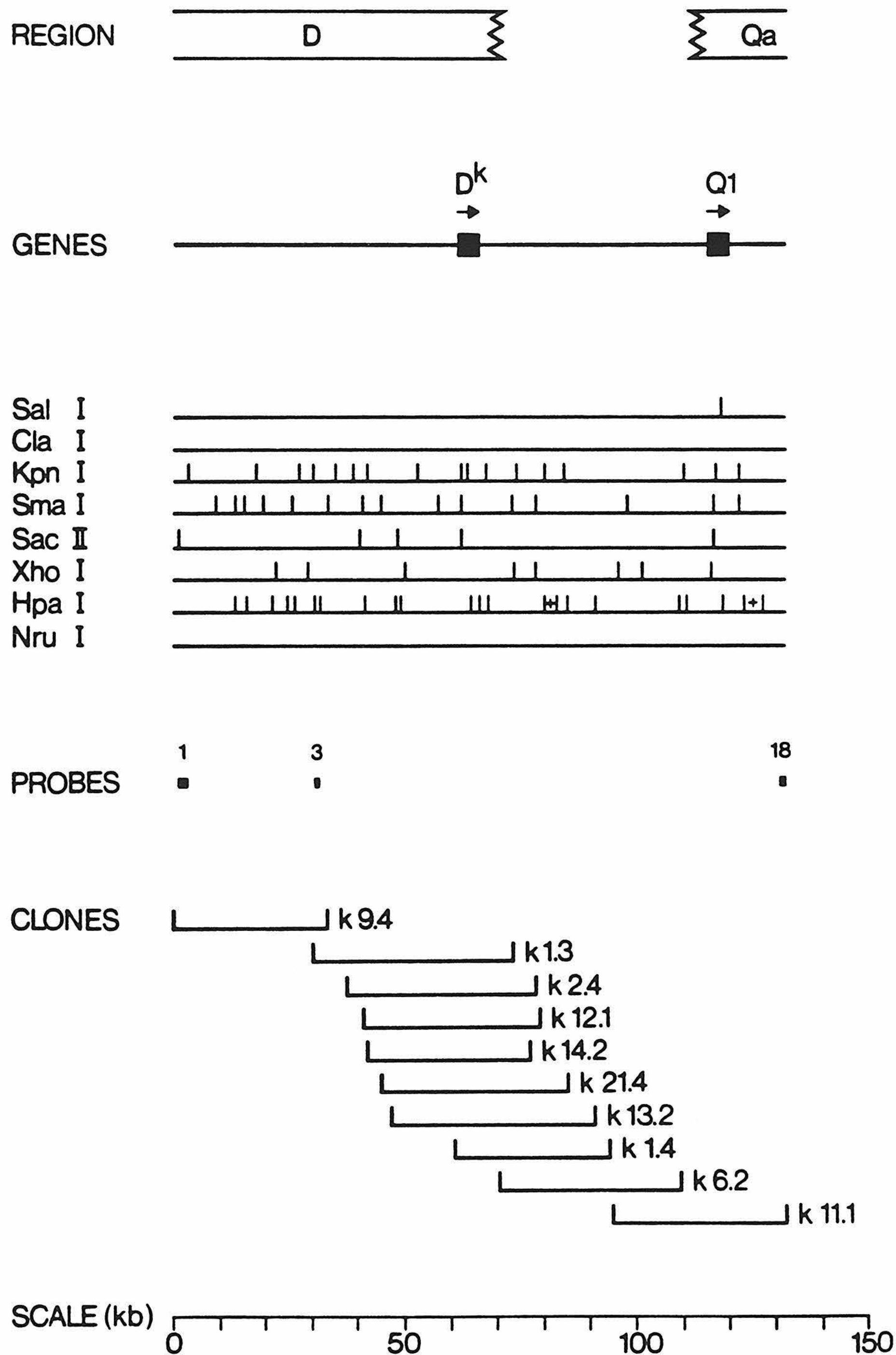

REGION
D
Qa
D^k
Q1
GENES
Sal I
Cla I
Kpn I
Sma I
Sac II
Xho I
Hpa I
Nru I
1
3
18
PROBES
CLONES
k 9.4
k 1.3
k 2.4
k 12.1
k 14.2
k 21.4
k 13.2
k 1.4
k 6.2
k 11.1
SCALE (kb)
0
50
100
150

Figure 4. Typing of D and Qa region specific sequences from AKR by Southern blot analysis of polymorphic restriction sites. Ten micrograms of DNA per lane were digested with the indicated restriction enzymes, run on a 0.6% agarose gel and blotted to nitrocellulose filters. Hybridization with the probes indicated was done as described (8).

A. Probe 1, isolated from cosmid clone k9.4 (compare Fig. 3 and Materials and Methods) identifies a polymorphic BamHI site in BALB/c and C3H mice. Analysis of the recombinant strains C3H.OH and A.AL maps the polymorphic site distal to the S region (compare Fig. 4D).

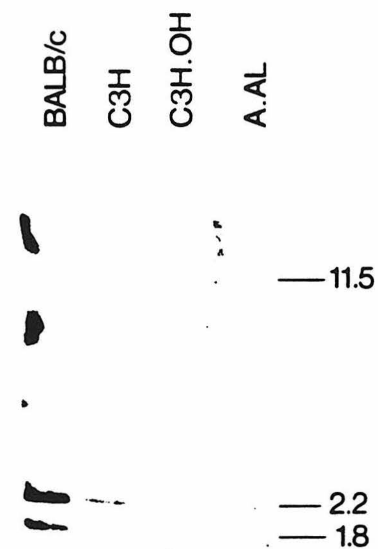
B. Probe 1 also identifies a polymorphic PstI restriction fragment in B6-H-2k and B6-Tla^{a} mice. Analysis of the recombinanats B6.K1 and B6.K2 maps the polymorphism proximal to the *Qa* region marker locus (compare Fig. 4E). Together with the mapping result from Fig. 4A, this maps probe 1 to the *D* region.

C. Probe 18 (compare Fig. 3 and Materials and Methods) identifies a 7.5 kb EcoRI fragment in B6-H-2^{k} which is missing in B6-Tla^{a}. Analysis of the recombinants B6.K1 and B6.K2 maps the polymorphism and therefore probe 18 to the *Qa* region (compare Fig. 4E).

D and E. *H-2* alleles in inbred mouse strains used (4). Vertical bars identify recombination events.

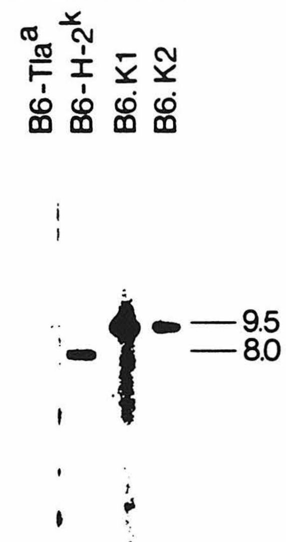
A. PROBE 1

Bam H I



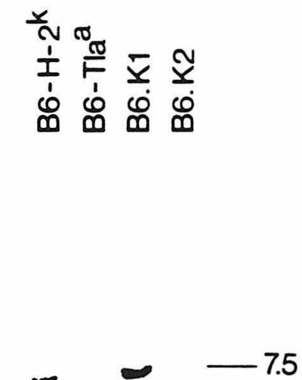
B. PROBE 1

Pst I



C. PROBE 18

Eco R I



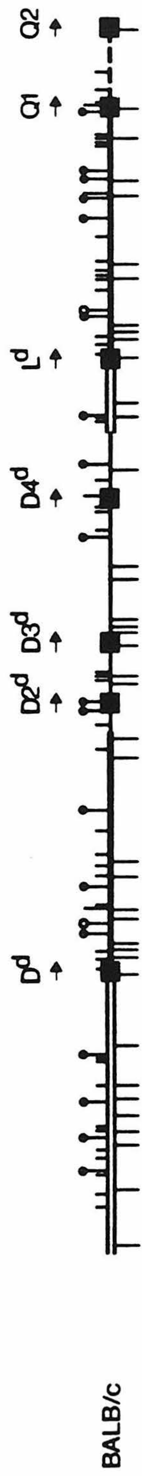
D.

	S		D	Qa
BALB/c	d		d	a
C3H	k		k	b
C3H.OH	d	I	k	b
A.AL	k	I	d	a

E.

	D		Qa		Tla
B6-Tla^a	b		a^{AorB6}		a^{AorB6}
B6-H-2^k	k		b^{AKR}		b^{AKR}
B6.K1	b	I	b^{AKR}		b^{AKR}
B6.K2	b		a^{AorB6}	I	b^{AKR}

Figure 5. Comparison of restriction maps for *D* and *Qa* region sequences of BALB/c, C57BL/10 and AKR mouse strains. Restriction sites indicated (SalI: , KpnI: , XhoI: , HpaI: , NruI:) were taken from Fig. 2 (BALB/c), Fig.3 (AKR) and Weiss *et al.* (28) (C57BL/10). Homologous segments based on restriction map similarity are shown by different symbols: 5' flanking sequences of D^d, L^d, D^b, D^k: open boxed line; 3' flanking sequences of D^d, L^d, D^b, D^k: thick line; 3' flanking sequences of *Q1*: dotted line.


BALB/c
Dd
D2d
D3d
D4d
Ld
Q1
Q2

C57BL/10
Db
Q1
Q2

AKR
Dk
Q1

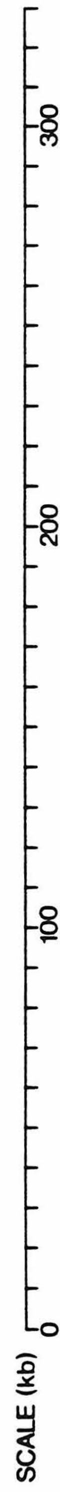
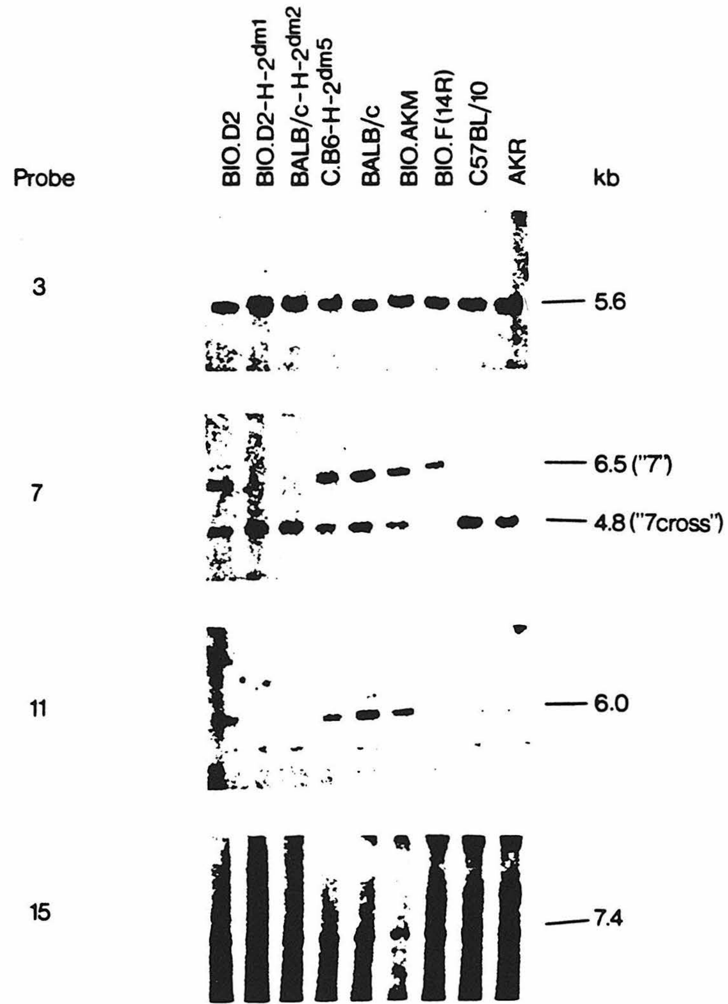

SCALE (kb)
0
100
200
300

Figure 6. Southern blot analysis of different inbred strains of mice and three d haplotype mutant strains with *D* region probes. Location of the probes is shown at the bottom (compare Fig. 2). The "7 cross" indicates a sequence in the 3' flank of L^d that cross hybridizes strongly to probe 7 (see also Fig. 7). 10 micrograms of DNA per lane were digested with BamHI and analyzed as described in the legend of Fig. 4. After hybridization the filters were washed twice for 20 min each at 68°C in 0.1X SSC, 0.1% SDS.

Gene expansion and contraction in the D region



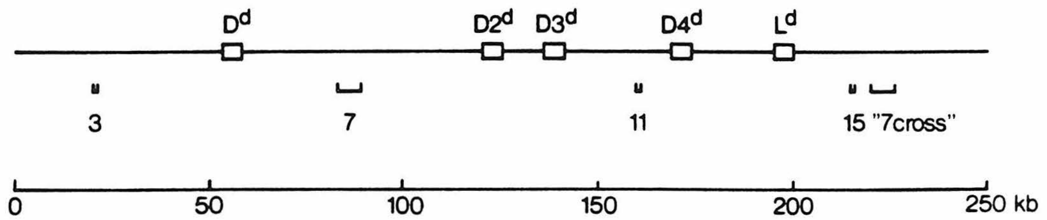


Figure 7. Cross hybridization patterns in the BALB/c *D* and *Qa* regions. Cross hybridization was checked on *K*, *D*, *Qa* and *Tla* region cosmids as described in the text. Hybridization probes indicated by filled boxes are the same in as Fig. 2. Cross hybridizing sequences are indicated by open boxes on the same line. Probes 1 and 3 showed no cross hybridization to any of the *K*, *D*, *Qa*, and *Tla* region cosmids tested under the stringency conditions used (washed at 0.5 x SSPE [45], 68°C, 1 hr). In addition to the sequences shown, probes 8 and 14 also gave strong cross hybridization to *K* region sequences and the following probes gave weak cross hybridization to: probe 4, 3' of D^d, 3' of *Qa*; probe 7, segments between *Q4*-*Q5*, *T3*-*T5*, *T6*-*T7*; probe 10, segments between *T3*-*T5*, *T6*-*T7*; probe 14, 3' of *Q1*, 3' of *Q2*.

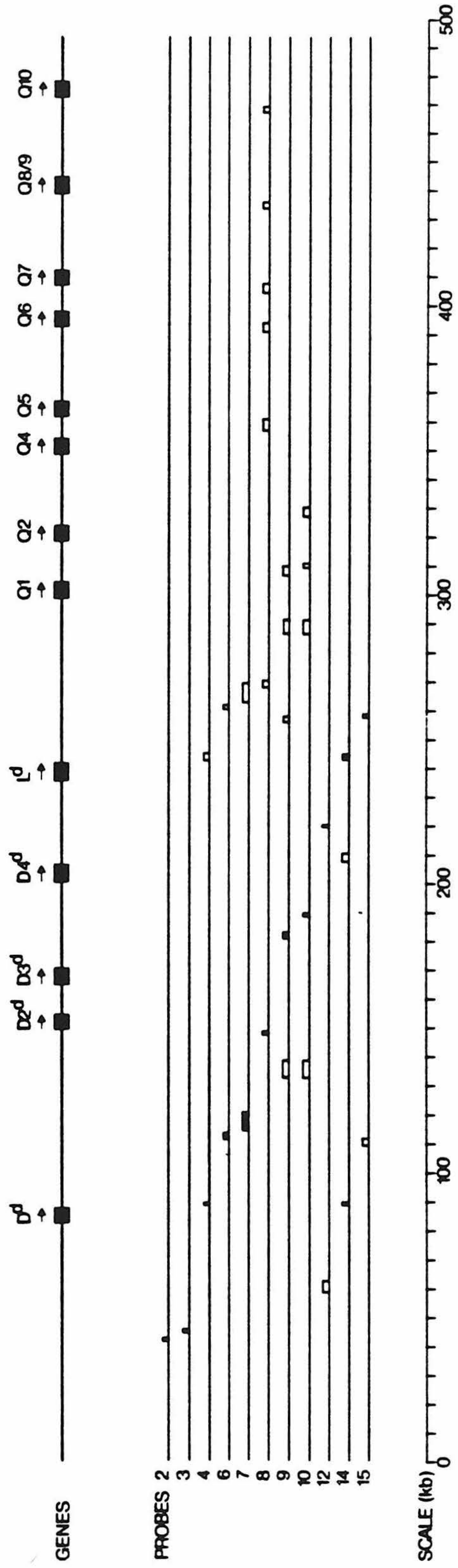

GENES
D^d
D2^d
D3^d
D4^d
L^d
Q1
Q2
Q4
Q5
Q6
Q7
Q8/9
Q10
PROBES 2
3
4
6
7
8
9
10
12
14
15
SCALE (kb)
0
100
200
300
400
500

Figure 8. Proposed evolutionary scheme for the generation of the five *D* region class I genes of the BALB/c mouse by unequal crossing over. The filled boxes denote the *D* and ***Qa*** genes on a BALB/c-like chromosome. The open boxes denote the allelic genes on a distantly related chromosome. A homologous recombination first exchanged the *D* genes. The resulting recombinant chromosome had later undergone an unequal crossing-over with the original BALB/c-like chromosome, generating an expanded chromosome.

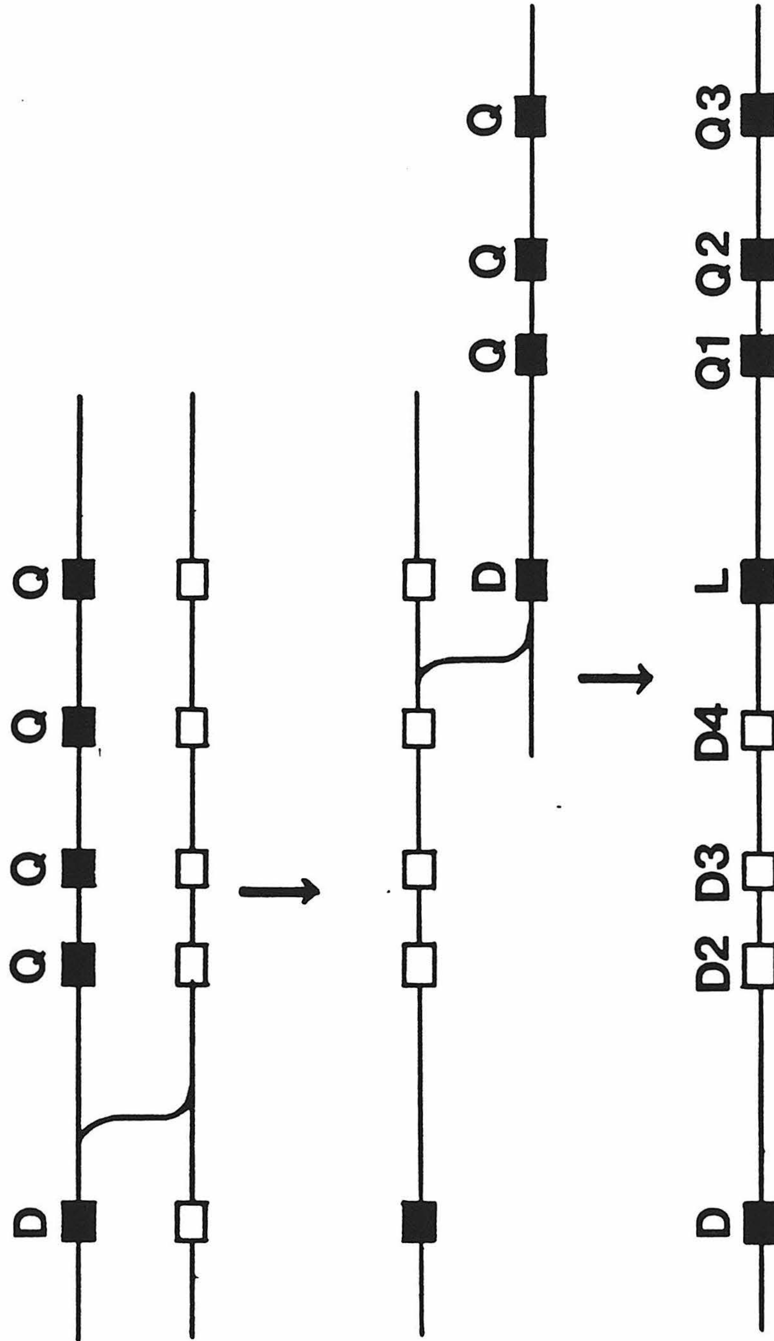

D
Q
Q
Q
Q
D
Q
Q
Q
D
D2
D3
D4
L
Q1
Q2
Q3

Table 1. AKR cosmids k12.1 and k13.2 encode the D^k class I molecule[a]

		Cells Tested [cpm bound]			
Antibody[b]	Specificity	IC9[c]	87-8-7.3[d]	87-1-6[d]	76D2.6[d]
15-5-5 S	D^k, K^d	556	7130	11528	8514
H11-3	D^k	986	7742	13958	9484
H142-23	D^k, K^k, K^b, K^s	498	5450	12264	8010
H100-27	D^k, K^k, r	908	3956	6172	4494
H100-30	D^k, K^k, s, b, r, q	942	4420	6648	4992
B22-249	D^b	6538	5348	6854	5048
K7-309	K^b	842	828	1220	1340
H116-22	K^k, t, q, s, r	1264	1470	1536	1228
R1-21.2	D^k, K^k, K^b, K^s, K^f, K^q	1072	7912	17108	13878
R28-27	I_i	1272	1308	1460	1980

[a]Data are from cellular radioimmunoassays. For mouse monoclonal antibodies ^{125}I-protein A was used; for rat monoclonals, first anti-rat κ chain and then ^{125}I-protein A was used.

[b]R1-21.2 is a rat anti-mouse H-2 antibody and R28-27 is a rat anti-mouse invariant chain antibody.

[c]IC9 is a (C57BL/6 x C3H)F_1 fibrosarcoma line expressing only D^b.

[d]87-8-7.3 and 87-1-6.3 are independent clones isolated from transfection of IC9 cells with plasmid pDk17 derived from cosmid k12.1. 76D2 is a clone selected from IC9 cells transfected with cosmid k13.2.

Table 2. AKR cosmids k1.3 and k12.1 encode the D^k target antigen

B10.CAS2 fibroblast transfected with cosmid	Exp. No.	Lysis by CTL[a]		Binding of monoclonal antibodies[b]		
		anti-D^k	anti-B10.CAS2	15-5-5 (anti-D^k)	16-1-2 (anti-D^k, D^{cas2})	18/20 (anti-TL)
k1.3	1	4.3	<u>11.3</u>	<u>275</u>	<u>1739</u>	40
	2	<u>16.6</u>	<u>10.7</u>	<u>1016</u>	<u>2155</u>	nd
	3	nd	nd	<u>263</u>	<u>2107</u>	25
k12.1	1	<u>20.7</u>	<u>23.5</u>	144	<u>765</u>	43
	2	<u>28.8</u>	<u>22.1</u>	438	436	nd
	3	nd	nd	160	<u>2105</u>	44
Controls[c]	1	-6.8 to -1.3	<u>14.6-31.8</u>	84-128	114-344	50-88
	2	-7.9 to 1.2	<u>11.7-39.6</u>	197-387	94	nd
	d	nd	nd	56-131	245-743	24-68

[a]Expressed as % of total ^{51}Cr released above control at 100 effectors per target cell. anti-D^k: (B10 x B10.CAS2)F_1 anti-B6.AK1; anti-B10.CAS2: C57BL/6 anti-B10.CAS2. Significant lysis underlined.

[b]cpm ^3H-leucine-labeled antibodies bound per 1-2 x 10^6 fibroblasts. Significant binding underlined.

[c]Transfected with L^d, D^d, or Tla region cosmids; from 1 to 12 per test.

nd = not done.

CHAPTER THREE

CLASS I GENES OF THE MURINE MAJOR HISTOCOMPATIBILITY COMPLEX ARE A PATCHWORK CREATED BY MULTIPLE RECOMBINATIONAL AND CONVERSIONAL EVENTS

CLASS I GENES OF THE MURINE MAJOR HISTOCOMPATIBILITY COMPLEX ARE A PATCHWORK CREATED BY MULTIPLE RECOMBINATIONAL AND CONVERSIONAL EVENTS

By Y. Henry Sun, Kurt Brorson, Leroy Hood, and Stephen W. Hunt III

From the Division of Biology, California Institute of Technology

Pasadena, California 91125

Introduction

The murine class I antigens are a closely related family of glycoproteins which include the transplantation antigens H-2K, H-2D and H-2L as well as the hematopoietic differentiation antigens Qa-1, Qa-2,3 and thymus leukemia (TL) antigen (for review, see Klein, 1979; Flaherty, 1981; Ploegh et al., 1981; Hood et al., 1983). Class I glycoproteins range in size from 34,000 to 45,000 daltons and associate noncovalently with β_2-microglobulin, a 12,000 dalton polypeptide. Class I molecules in general are integral membrane proteins, although in at least one case the protein product is secreted (Kress et al., 1983). Class I antigens are composed of three external regions, each approximately 90 amino acids in length, a hydrophobic transmembrane domain of about 40 residues and a cytoplasmic domain of approximately 30 residues (Coligan et al., 1981). This family of proteins can be divided into two subfamilies based on their patterns of expression and their function. The transplantation antigens are expressed on most, if not all, somatic cells. These molecules are extremely polymorphic and are involved in the lysis of virus-infected and neoplastically transformed cells by cytotoxic T lymphocytes (Zinkernagel and Doherty, 1979). The other group of proteins has been called the hematopoietic differentiation antigens because they have a much more restricted tissue distribution which allowed them to be used as differentiation markers. They are less polymorphic (Tewarson et al., 1983) and their function, if any, is unknown (Yokoyama et al., 1983).

The class I genes reside in the major histocompatibility complex (MHC), located on chromosome 17 (Lader et al., 1985) of the mouse. The MHC spans about 2 centimorgans (approximately 4000 kilobases) of DNA and is divided into six genetic regions denoted *K*, *I*, *S*, *D*, *Qa* and *Tla*. The BALB/c mouse contains at least 34 class I genes (Figure 1). (Steinmetz et al., 1982; Rogers, 1985; Fisher et al., 1985; Steinmetz et al., 1986). These genes have been cloned and their

locations within the MHC have been determined (Winoto et al., 1983). There are two class I genes in the *K* region, five in the *D* region, eight in the *Qa* region and 19 in the *Tla* region. The genes encoding the class I antigens are each split into eight exons, corresponding to the structural components of the protein (see Hood et al., 1985). The first exon encodes the leader or signal peptide; exons 2, 3 and 4 encode the three external regions; exon 5 encodes the transmembrane domain; and exons 6, 7 and 8 together encode the cytoplasmic domain.

Several investigators have studied the evolutionary process that has generated the large number of class I genes. Rogers (1985) subcloned and sequenced the region containing exons 2 and 3 from 15 of the BALB/c class I genes. Using these cloned regions as hybridization probes, Rogers classified the class I genes into families based on homology with the probes. However, the regions examined have been shown to be functionally important in that the allo-antigenic determinants recognized by antibodies and cytotoxic T cells are encoded by them (see Stroynowski et al., 1985). Consequently, selection may obscure the true evolutionary history of these genes. Steinmetz et al. (1982, 1984) and Weiss et al. (1984) based their comparisons on the pattern of restriction sites, which are relatively uniformly distributed throughout the entire region and presumably are free of functional selection. However, the amount of sequence data actually compared is very limited due to the short recognition sequences of restriction enzymes.

In this report, we describe the sequencing of the transmembrane exons of most of the class I genes by a general method. Comparison of the DNA sequences led to the identification of unique stretches from which a set of highly specific oligonucleotide probes were made. The transmembrane probes and a set of 23 noncoding flanking region probes were hybridized to cosmid clones spanning the *K*, *D*, *Qa* and *Tla* regions of the MHC. Analysis of the compiled transmembrane

sequences as well as the coding and noncoding sequence hybridization data has allowed a unique opportunity to study the interrelationships and evolution of the genes in the BALB/c class I gene family. The results indicate that there have been numerous events leading to the present state of the class I gene family. These events include point mutations, duplications, insertions, deletions, inversions, and gene conversions.

RESULTS

DNA Sequencing of the Transmembrane Exon of 20 Class I Genes

To study the relationships between class I genes, it is desirable to compare the entire genomic sequences of all 34 genes. Since this would require sequencing between 170 and 270 kb of DNA, it is not yet feasible. The best alternative is to compare sequences taken from the same region of all 34 genes. The transmembrane exon was chosen as the region to be sequenced for two reasons. First, the transmembrane exon was among the most variable exons in the class I genes (Fisher et al., 1985). Presumably, the major constraint on the divergence of this exon is the need to maintain hydrophobicity in the transmembrane domain. Second, and more importantly, the transmembrane exon was located approximately 120 nucleotides downstream of the most highly conserved exon in class I genes. Exon 4, which encodes the β_2-microglobulin-binding domain, exhibits between 90-95% homology among all of the sequenced class I genes (Fisher et al., 1985). As a result of this fortuitous coincidence, we have devised a general procedure to sequence the transmembrane exon for most of the class I genes (Figure 2). An invariant 18-nucleotide sequence was identified in the fourth exon of several class I genes. This sequence is located approximately 70 nucleotides 5' to the end of the fourth exon. Using an oligonucleotide derived from this sequence as primer, we were able to sequence through the end of the fourth exon,

across the fourth intron (120-140 bp), and into the transmembrane exon in a single sequencing reaction using the dideoxynucleotide chain termination method (Sanger et al., 1977).

Twenty transmembrane DNA sequences were determined in this study. DNA sequences for eight other genes, namely, $H\text{-}2L^d$ (Moore et al., 1982; Evans et al., 1982), $H\text{-}2K^d$ (Kvist et al., 1983), $H\text{-}2D^d$ (Sher et al., 1985), Q7 (Steinmetz et al., 1981), $T1^c$ (Fisher and Hood, in preparation), $T13^c$ (Fisher et al., 1985), $T10^c$ (Hunt and Robinson, in preparation) and "37" (Lalanne et al., 1985) have previously been determined. In addition, sequences of the Q10 and T3 genes from the C57BL/10 mouse (Mellor et al., 1984; Pontarotti et al., 1986) were included in the analysis. The Q10 gene appears to be highly conserved among different strains of mice. The C57BL/10 T3 gene has been shown to be highly homologous to the BALB/c T3 gene (Fisher et al., 1985; Pontarotti et al., 1986). The D3 transmembrane sequence was unobtainable using this generalized method because it apparently lacks a conserved fourth exon (data not shown). Altogether, a total of 30 transmembrane sequences were used in the following comparison.

Comparison of Transmembrane DNA Sequences Identifies Gene Subfamilies

The DNA sequences of the 30 transmembrane exons plus their 5' and part of the 3' introns are shown in Figure 3. Pair-wise homology comparisons between the transmembrane exon sequences are shown in Table 1. Several pairs of genes, D/Q5, Q4/Q6, Q7/Q8/9, T3/T13, T2/T12 and T9/T17, are 96-99% homologous and, as will be discussed later, are probably the result of recent duplication events. The T4 and T6 genes form another pair with 97% homology in the 5' intron and 93% homology over the first 46 nucleotides of the exon. The T6 gene then diverges with a long poly(T) stretch. The remainder of its transmembrane exon was not sequenced.

Six subfamilies can be identified based on the degree of homology. Group 1

includes the *K*, K2, *D*, Q5, D2, D4, *L*, Q1, Q2, Q4, Q6, Q7, Q8/9, Q10, T4, T6, T8, T16 and "37" genes. These genes are about 80-90% homologous to each other, with the exception of Q10, which is only about 70-80% to the others. Q1 has an 18-nucleotide insertion which is 78% homologous to the preceding 18 nucleotides and 94% homologous to the corresponding stretch in T4. This insertion may have resulted from a duplication or unequal crossover event. Group 2 includes a single member, T10, which, although related to group 1, was suitably different so that it could be assigned to a separate group. As discussed in the next section, the T18 gene probably also belongs to this group. Group 3 includes the gene T1 and the T3/T13 gene pair, with 87-88% homology. Gene T11 is probably also a member of this family. All four of these genes hybridize to a genomic fragment containing the transmembrane exon derived from the T13 gene (Fisher et al., 1985). T11 also hybridizes to an oligonucleotide probe derived from the T1 transmembrane sequences (data not shown). Group 4 includes the gene T5 and the T2/T12 gene pair. T5 is 91% and 95% homologous to T2 and T12, respectively, from the beginning of the exon down to position 295 and then diverges greatly, possibly as a consequence of a recombinational event. Groups 5 and 6 include the gene pairs T7/T15 and T9/T17, respectively. Between groups, there is about 30-60% homology.

Among members of the same group, the homologies at the 5' and 3' intron region are generally higher than within the transmembrane exons (analysis not shown), and will be discussed later. Between groups, however, the homologies are high at the 5' intron (about 60-80%) but rapidly drop off after the 5' intron. This gradient of homology is probably the residual effect of conservation of the fourth exon, which encodes the highly conserved β_2-microglobulin-binding domain.

Most of the Exon Sequences Encode a Functional Transmembrane Protein Segment

The translation of the DNA sequences of the transmembrane exons are

shown in Figure 3. Five of these genes, *K*, *D*, *L*, T3 and T13, are known to encode membrane-bound proteins (Goodenow et al., 1982a; 1982b; Fisher et al., 1985; Pontarotti et al., 1986). Genes K2, Q5, D2, D4, Q1, Q2, T4, T8, T16, "37", T2, T9 and T17 all have a similar apparently functional transmembrane segment, namely, a long stretch of mostly noncharged amino acids ending with several positively charged residues. Q1 has an 18-nucleotide insertion which resulted in six extra amino acid residues. The extra length would probably have little effect on the stability of the protein in the membrane. Q2, although having a typical transmembrane segment, does not have a fourth exon as determined by the lack of hybridization to a probe containing the entire fourth exon from the $H\text{-}2D^d$ gene (data not shown). T1 has three positively charged residues in the membrane-spanning region; it is not known whether this segment can be membrane-bound.

Several sequences terminate prematurely due to single nucleotide deletions which cause a shift in the translation reading frame and lead to in-phase termination codons. The translated Q7, Q8/9 and T10 transmembrane segments terminate shortly before the end of the exon, but are still 35 residues long and end with two charged residues. Studies have shown that class I molecules lacking the cytoplasmic domain and the cytoplasmic-side of the transmembrane region, including the charged amino acid residues, can still be membrane-bound (Zuniga et al., 1983; Zuniga and Hood, 1986). Indeed, the transmembrane exon of Q7 has been shown to be capable of encoding a membrane-bound segment (Stroynowski et al., 1984). $Q10^b$ encodes a secreted class I molecule (Kress et al., 1983; Maloy et al, 1984; Devlin et al., 1985). Q4 and Q6 are also likely to be secreted; their last 10 amino acid residues are 80% homologous to the last 10 residues of the $Q10^b$ transmembrane segment. T5 and T12 encode even shorter protein segments, which would most likely cause these proteins to be secreted.

T7 and T15 have no positively charged amino acids at the end of the exon. It is not clear whether such proteins would be membrane-bound or secreted.

In summary, 19 genes apparently encode functional transmembrane segments, and 8 genes are likely to encode a secreted protein.

Gene-specific Oligonucleotide Probes

By comparison of the transmembrane exon sequences, stretches of 16-21 nucleotides which differed from the other genes by two or more nucleotides were selected, and a set of synthetic oligonucleotides was made. The sequences used for the oligonucleotides are underlined in Figure 3. The oligonucleotides were actually synthesized using the antisense strand sequences such that the probes could be used for hybridization to both DNA and RNA blots. As predicted from the sequences, these oligonucleotides are highly specific as hybridization probes (data not shown). Under the most stringent conditions used in this study (0.5X SSC, 50°C), these probes may cross-hybridize to sequences with a single nucleotide mismatch (data not shown). With more stringent washes these probes are expected to be gene-specific. These gene-specific hybridization probes will be extremely useful in the study of this large multigene family, especially in distinguishing the expression of individual genes.

The T1 and T10 probes hybridize to the unsequenced T11 and T18 genes, respectively, indicating that T11 and T18 probably belong in groups 3 and 2, respectively. The T1 probe also cross-hybridizes to a region about 15 kb 5' to the T1 gene (Figure 5), and will be discussed later. The T4 probe cross-hybridizes, under less stringent conditions, with T6, T8, T16 and the unsequenced T14 gene, indicating that the T14 gene belongs in group 1. Hybridization with the T9 probe allowed us to determine the transcriptional orientation of the T9 and T17 genes (Figure 5).

Hybridization with Flanking Sequence Probes

We have used 23 low copy-number sequences isolated from regions flanking the various class I genes (Figure 5; see Materials and Methods) as hybridization probes to analyze the relationships of the class I genes. Fourteen probes have been isolated from the *H-2D* region, three from the *Qa-2,3* region, five from the *Tla* region, and one from the *H-2K* region. These probes were radioactively labeled and hybridized to DNA dot blots (data not shown) containing 33 cosmid clones covering 1000 kb and representing the entire cloned class I region. The clones that hybridized to each probe were then digested with several restriction enzymes and analyzed by Southern blots to locate the region where each probe hybridized (Figure 5).

Several interesting points can be made from this analysis. First, certain patterns repeated themselves (Figure 5). The most striking similarity is between the region flanking the D^d gene (position 50-130 kb) and the region flanking the L^d gene (position 210-300 kb). The relative positions of 19 fragments cross-hybridizing to 15 probes are remarkably conserved. Although comparison of restriction site maps have shown homology between the regions immediately flanking D^d and L^d (Steinmetz et al., 1982), this large homology unit was not apparent. The extensive homology between the flanking regions of the D^d and L^d genes strongly supports the hypothesis that these two genes are the result of a very recent duplication event. Equally striking is the repeating pattern of five probes (5, 6, 15, 16 and 17) occurring around the Q4-Q10 genes, and of four probes (7, 8, 9 and 12) occurring around the Q1 and Q2 genes. These repeating patterns again suggest that these genes arose by duplication.

In addition to the duplications noted above, processes such as insertion, deletion and inversion can be shown to have occurred within this region. For example, the distance between the fragments hybridizing to probes 11 and 17 in

the 5' flanking regions of the D^d and L^d genes is 15 kb and 9.5 kb, respectively. Presumably, a 5.5 kb insertion in the 5' flanking region of the D^d gene, or a deletion in the 5' flanking region of the L^d gene, has occurred after the duplication of the two genes. An example of inversion is shown by the relative position of probes 6 and 17 in the 5' flanking region of the Q8/9 and Q10 genes. The order is 15-16-6-17 for Q8/9 and 15-16-17-6 for Q10. Presumably, the 10 kb region containing probes 6 and 17 has undergone an inversion in the 5' flanking region of either the Q8/9 or Q10 genes.

The order of probes 21 and 22 in the 5' flanking region of T1 is inverted relative to that in the 5' flanking region of the T11 gene. Weak hybridization to a class I 5' coding sequence probe and to the T1 transmembrane oligonucleotide probe has been noted in the flanking region 5' to the T1 gene (Steinmetz et al., 1982; Figure 5), suggesting that both a duplication of a T1-like gene and an inversion of the duplicated gene have occurred.

Although the K^d, D^d and L^d genes are about equally homologous to each other (Table 1; Fisher et al., 1985), the flanking region of the K^d gene is significantly different from those of the D^d and L^d genes, which are themselves highly homologous as discussed above. This observation suggests that the K^d gene was derived from a different ancestral gene than the one which generated the D^d and L^d gene pair, and that it was later modified by gene conversion events. In fact, the flanking regions of *K* and K2 resemble the flanking regions of genes in the *Qa* region, suggesting that they were derived from the *Qa* region. Similarly, while genes *D* and Q5 have almost identical transmembrane exons, there is no apparent homology at their flanking regions, and an oligonucleotide probe derived from the third exon of the D^d gene does not hybridize to the Q5 gene. The gene pairs *K* and D2, Q1 and Q4, T4 and T6, have 95-97% homology in the 5' intron, but show no apparent similarity in their flanking regions. These data are suggestive of gene conversions or multiple recombinations.

All of the probes, except 1, 2 and 10, hybridize to more than one fragment in the class I region (Figure 5). Since the probes were selected for single or low copy-number, there cannot be a bias for duplicated sequences. This suggested that either all of the DNA rearrangement and duplication events involved only the sequences within the MHC, or that the region undergoes very rapid duplications such that any exogenous sequences brought in are quickly duplicated. Probes 1 and 2 are about 30 to 40 kb 5' to the D^d gene, and perhaps reside outside of this region of rapid evolution.

DISCUSSIONS

Qa and *Tla* Region Genes Have Been Duplicated

The comparison of the transmembrane sequences and the pattern of hybridization with the flanking sequence probes together present compelling evidence that there have been duplications of large blocks of DNA in the *Tla* region of the BALB/c mouse as suggested previously (Rogers, 1985; Fisher et al., 1985; Hammerling et al., 1985). It is apparent that a block of DNA containing genes T1, T2 and T3 has been duplicated to yield the T11, T12 and T13 genes. A second block of DNA containing genes T6, T7, T8 and T9 was duplicated to yield the T14, T15, T16 and T17 genes. It is possible that the T10 gene was also included in this duplication event and yielded the "37" gene. Although the T10 transmembrane exon sequence has diverged significantly from "37" and the other members of group 1, a fragment isolated from the "37" cDNA clone (Lalanne et al., 1985) cross-hybridizes to the T10 gene. A genomic clone containing the "37" gene has recently been isolated (C. Transy, personal communication). It will be interesting to determine whether the "37" gene resides in the Tla region.

These duplications must have occurred after the separation of the BALB/c and the C57BL/10 inbred strains. The C57BL/10 mouse has a single large cluster

of *Tla* region genes, the restriction map of which is highly similar to the T1-T10 region of the BALB/c mouse (Weiss et al., 1984; Fisher et al., 1985), and is likely to represent the ancestral configuration. The T11-T13 region and the T14-T17 region could have duplicated independently, or the entire T1-T10 region was duplicated and genes T4 and T5 were subsequently deleted.

Duplication events have also been proposed in the *Qa* region (Weiss et al., 1984; Mellor et al., 1985; Devlin et al., 1986). The present data lend further support to the hypothesis that the Q4-Q10 genes have been duplicated in pairs. The flanking regions of Q4-Q10 are highly similar. The Q6-Q7 region (map position 383-427 in Figure 5) probably represents the configuration of the ancestral gene pair. Duplications resulted in the Q4/Q5, Q6/Q7, and Q8/Q9 pairs. The three gene pairs in the B10 mouse are highly similar (Weiss et al., 1984). By restriction map comparison with the *Qa* genes in the C57BL/10 mouse, the BALB/c Q8/9 gene has been suggested to be derived from a fusion of the original Q8 and Q9 gene (Weiss et al., 1984). Similarly, the region between the Q4 and Q5 genes has undergone a deletion, bringing closer the distance between the two genes. A gap between these two genes has been shown in comparison to the C57BL/10 Q4-Q5 region (Weiss et al., 1984). The suggestion that these genes were duplicated in pairs would suggest that the Q4, Q6 and Q8 genes would be homologous to each other, and the Q5, Q7 and Q9 genes would be homologous to each other. The transmembrane sequences show that Q4 and Q6 are almost identical as are Q7 and Q8/9, the 3' half of which presumably came from the Q9 gene. The Q5 gene, however, has diverged significantly from Q7 and Q8/9, both in the 5' half of the gene (Rogers, 1985) and in the transmembrane exon.

Probes 7, 8, 9 and 12 were repeated three times at map position 267-328 (Figure 5), indicating that the Q1, Q2 and Q3 genes were also duplicated from a common ancestral gene. The Q3 gene has been deleted in the BALB/c mouse

genome, but is still present in the C57BL/10 mouse (Weiss et al., 1984).

The *H-2D* and *H-2L* genes and their flanking regions are highly homologous and must be the result of a very recent duplication which occurred after the separation of the BALB/c and B10 mouse strains. In fact, the D^b gene is more homologous to the L^d gene than to the D^d gene, and some have suggested it is allelic to L^d (Reyes and Wallace, 1982). Since the C57BL/10 mouse contains only a single *D* gene (D^b), the BALB/c L^d gene was probably its allele and has duplicated to form the D^d gene. The D2-D4 genes are more distantly related to the *D*, *L* and *Qa* genes. Their possible origin has been discussed (see Chapter Two of this thesis).

The duplications discussed above probably resulted from homologous but unequal crossovers.

Comparisons with Flanking Region, 5' Exon, and Transmembrane Probes Indicate the Class I Genes Are a Patchwork Created by Multiple Recombinational and Conversional Events

Many examples have already been given above that multiple events such as duplication, insertion, deletion, inversion and gene conversion have occurred in the class I regions of MHC. Incorporation of the hybridization data using probes generated from the 5' half of the genes (Rogers, 1985) further demonstrates that the class I genes are a patchwork of multiple events.

The flanking region of T5 is homologous to that of the T4 gene and the 5'-half gene between T3 and T4. However, the T5 transmembrane sequence is very homologous to those of the T2/T12 gene pair, while the second and third exons of T5 are homologous to those of the T15 gene (Rogers, 1985). The second and third exons of the T15 gene, in turn, also show homology to the K gene (Rogers, 1985). Examples like this are numerous and will not all be described here. This patch-work pattern (i.e., a part of the gene is homologous to one set of genes while

another part of the gene is homologous to a second set of genes) is probably the result of multiple gene conversion or recombination events. Attempts to construct a phylogenetic lineage of these genes based on any single set of information is likely to be misleading.

Comparison of the transmembrane sequences have shown that the genes in the large group (group 1) are approximately equally homologous to each other. Members of this group include genes from the *H-2K*, *H-2D*, *Qa* and *Tla* regions, and analyses with flanking region probes have shown that they are not all closely related. This observation suggests that there has been interaction between these genes to create the homogeneity. Careful analysis indicates that clustered non-random changes occur in the exon, while the differences in the introns are more uniformly scattered (Figure 3). These clustered changes suggest that the interaction probably occurred through the action of gene conversion-type of mechanism (Jaulin et al., 1985; see also Hood et al., 1983). These clustered changes in the exon also caused the apparent lower homology in the exons than in the introns. Furthermore, no single donor or acceptor of the gene conversions can be identified, suggesting that all of these genes can serve both as donor and acceptor. Previous analysis of the complete sequences of the K^d, K^b and L^d genes (Weiss et al., 1983) has failed to identify the clustered changes in the transmembrane exon, probably due to the small number of sequences compared. The occurrence of the CpG dinucleotide is related to the clustered changes (Figure 3) as previously reported, and was suggested to increase the frequency of double strand break repairs and hence the frequency of unequal repairs (Jaulin et al., 1985). In this respect, it is interesting to note that the frequency of the CpG dinucleotide is very low (26/5762) in the transmembrane exon and its flanking introns (Figure 3).

How Many *Qa* and *Tla* Region Genes are Expressed? How Many Are Essential?

Although there are at least 34 class I genes in the BALB/c mouse, only a few class I antigens have been serologically or biochemically defined. It has been suggested that the large number of genes may simply serve as a repertoire responsible for the generation of polymorphism in the H-2 class I antigen, and hence many of these genes may be pseudogenes. Alternatively, the genes may encode bona fide class I proteins that are expressed but have not yet been identified because of the lack of suitable serological probes. The inability to generate antibodies against these proteins may be because these proteins are expressed at very low levels, are expressed only on a small population of cells or are expressed only at a specific developmental stage. In addition, these proteins may not exhibit enough polymorphism between different mouse strains to allow the production of alloantisera.

Several genes have definitely been identified as pseudogenes. The D3, T1, T4 and T5 genes have been shown to be pseudogenes on the basis of frameshift or nonsense mutations in the second or third exons (Rogers, 1985; Fisher and Hood, in preparation). Since the T11 gene was probably duplicated from the T1 gene, it is also likely to be nonfunctional. The D3 and Q2 genes do not contain a region which hybridizes to an *H-2D*^{d} fourth exon probe. These two genes either do not contain a fourth exon or their fourth exons have diverged so greatly that they would probably encode an $\alpha 3$ domain unable to bind to β_2-microglobulin. Therefore, if a protein product is made at all, it would not resemble a typical class I antigen. There is no evidence that any of the remaining 28 genes are nonfunctional.

The *K*, *D*, *L*, and T13 genes have been shown to be expressed (Goodenow et al., 1982a; 1982b). The Q6, Q7, Q8 and Q9 genes from the C57BL/10 mouse have been shown to encode Qa-2/3-reactive polypeptides, while no Qa-2/3-reactive polypeptides can be detected from the Q1-Q5 genes (Mellor et al., 1985). cDNAs

derived from the Q7, Q10 and "37" genes have been identified in the liver of DBA/2 mice (Lalanne et al., 1985), which, as far as can be determined by serology and by genomic Southern blot analysis with several probes, has an MHC indistinguishable from that of the BALB/c mouse. The Q10 gene in the C57BL/10 mouse is transcribed in the liver and encodes a secreted protein (Kress et al., 1983; Maloy et al., 1984). A partial cDNA transcript of the $T10^{C}$ gene has been identified in SV40-transformed NIH3T3 cells (Hunt, and Robinson, in preparation). Finally, on the basis of hybridization with the transmembrane oligonucleotide probes, three additional genes, D4, Q4 or Q6, and T9 or T17, have been shown to be expressed in the thymus of BALB/c mice (Hunt, in preparation).

Some of these genes are not essential for the survival of the animal, at least under laboratory conditions. Studies of two mutants, dm1 and dm2, indicate that the D2, D3, and D4 genes can be deleted, and that either of the D or L genes can also be deleted (Sun et al., 1985; Chapter Two) without deleterious effect on the mice. The fact that the C57BL/10 mouse does not have the D, D2-D4, and T11-T18 genes also indicates that these genes are redundant. Genes T1, T4 and T5 are pseudogenes and therefore are obviously not required. The observation that the BALB/cBy substrain has two gene fusion events (Q6/7 and Q8/9) (Mellor et al., 1985) and is phenotypically $Qa\text{-}2^{-}$ indicate that these genes, although encoding the Qa-2/3 polypeptides, are also not essential. Although these genes may still have a role in acting as a reserve of polymorphism for the H-2 class I genes, they clearly cannot have an essential role in the development of the animal.

MATERIALS AND METHODS

Cosmid Clones and Dot Blots

DNA of 34 cosmid clones spanning the entire cloned class I region were used in the dot blots. These representative cosmid clones are: *H-2K* region, II6.4,

II5.8, 17.1, II6.3 and II6.13; *H-2D/Qa* region, HS30.5, II4.12, HS17.1, II2.20, 4.1, 50.2, 5.4, JS4.4, 59.2, JS12.1, II3.20, 2.1, 64.1, 65.1, 12.1, 38.1, 30.1, 46.1 and 15.1; *Tla* region, 66.1, 59.1, 49.1, 5.2, 59.3, 12.2, 22.1, 6.3, 47.1 and 15.3. These clones have been described previously (Steinmetz et al., 1982; Steinmetz et al., 1984; Fisher et al., 1985; Chapter Two).

Each clone was first checked by comparing its digest with the restriction map. The DNA concentrations were measured by $O.D._{260}$ and then adjusted by the intensity of ethidium bromide staining of the digests run on agarose gel. 50 ng DNA of each clone was denatured with 0.5 N NaOH at 80°C for 10 min, neutralized by adding equal volume of 2M NH_4Ac, and filtered through Gelman Biotrace RP nylon membrane (gives stronger signals than Schleicher & Schuell BA85 nitrocellulose, NEN Gene Screen and ICN Biotrans nylon membranes) prewetted with 1 M NH_4Ac, using a Schleicher & Schuell manifold. The filters were washed in 2X SSPE and baked at 80°C under vacuum for 1-2 hours.

Sequencing of the Transmembrane Exons

DNA of cosmid clones containing the various class I genes was digested with BamHI, Sau3A or SmaI and shotgun-cloned into the BamHI or SmaI site of M13mp18. The recombinant phages were screened by hybridization to an oligonucleotide probe, LD1933. The probe has the invariant fourth exon sequence (see text) 5'ACCTTCCAGAAGTGGGCA 3' and will hybridize only to the recombinant phages with the desired insert in the proper orientation for sequencing toward the 3' end of the gene.

DNA sequencing was performed as described previously (Strauss et al., 1986) using the LD1933 oligonucleotide as primer.

Oligonucleotide Probes

Oligonucleotides were synthesized and purified as described previously (Strauss et al., 1986). They were radioactively labeled with γ-P^{32}-ATP by T4

polynucleotide kinase and 10^5 cpm/ml were used in each hybridization. Hybridizations were as described (Sun et al., 1985) except that they were at 37°C for 6-12 hr. Washes were at increasing stringency: 5X SSC at 37°C; 2X SSC at 37°C; 2X SSC at 43°C; 2X SSC at 50°C; 0.5X SSC at 50°C. All washing buffers contain 0.1% SDS. Autoradiographies were performed at -70°C using Kodak X-Omat XAR-5 film with Dupont Cronex intensifying screens.

Flanking Region Probes

For convenience, the probes used in this study are numbered 1 to 23. They correspond to: 1, dm1-38.1; 2, 1.3^k; 3, 18.1B; 4, II2.20.34; 5, II2.20wp; 6, II6.21.3wp; 7, 50.2; 8, 4.4.5; 9, 16.1.3; 10, 16.1A; 11, 59.2B; 12, 59.2A; 13, II1.10; 14, dm1-7.2; 15, 46.1; 16, 36.2S; 17, 36.2L; 18, 17.1; 19, 49.1S; 20, 49.1L; 21, 12.2; 22, 22.1; 23, 47.1. Probes 1, 2, 3, 4, 5, 7, 8, 9, 11 and 13 correspond to probes 2, 3, 4, 6, 7, 9, 10, 11, 12 and 15, respectively, previously described (Chapter One; Chapter Two). Probe 12, 16-23 have been previously described (Winoto et al., 1983). Probe 6 is a 0.5 kb SmaI/HpaI fragment from cosmid clone II6.21 subcloned into pUC8. Probe 10 is a 0.6 kb Sau3A fragment of a 4.7 kb HpaI fragment of cosmid clone 16.1 (Steinmetz et al., 1982) subcloned into the BamHI site of M13mp8. Probe 14 is a 0.8 kb Sau3A fragment of a 5.2 XhoI fragment of cosmid clone dm1-7.2 (Sun et al., 1985) subcloned into the BamHI site of M13mp8. Probe 15 is a 3.0 kb ClaI/SmaI fragment of cosmid clone 46.1 (Steinmetz et al., 1982) subcloned into the SmaI site of M13mp8. Probes 10 and 15 are generous gifts from Astar Winoto (California Institute of Technology, Pasadena, CA).

The insert of each M13 clone was cut out with EcoRI and PstI and gel purified. Four clones, 6.21.3, 36.2, 49.1 and dm1-7.2, generated two fragments which were purified separately and labeled L and S, respectively, to indicate their relative size.

The purified fragments were labeled according to the method of Feinberg and Vogelstein (1984) to specific activities of 1-4 x 10^6 cpm/ug. Hybridizations were as described (Sun et al., 1985). Washes were at 0.5X SSPE, 68°C for 1 hr, which would allow detection of sequences with greater than 90% homology to the probe.

REFERENCES

Coligan, J. E., Kindt, T. J., Uehara, H., Martinko, J., and Nathenson, S. G. (1981). Primary structure of a murine transplantation antigen. Nature **291**, 35.

Devlin, J. J., Lew, A. M., Flavell, R. A., and Coligan, J. E. (1985). Secretion of a soluble class I molecule encoded by the *Q10* gene of the C57BL/10 mouse. EMBO J. **4**, 368-374.

Devlin, J. J., Weiss, E. H., Paulson, M., and Flavell, R. A. (1986). Duplicated gene pairs and alleles of class I genes in the *Qa2* region of the murine major histocompatibility complex: A comparison. EMBO J. in press.

Evans, G. A., Margulies, D. H., Camerini-Otero, R. D., Ozato, K., and Seidman, J. G. (1982). Structure and expression of a mouse major histocompatibility antigen gene, $H\text{-}2L^d$. Proc. Natl. Acad. Sci. USA **79**, 1994-1998.

Feinberg, A. P., and Vogelstein, B. (1984). Addendum: A technique for radio-labeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **137**, 266-267.

Fisher, D. A., Hunt, S.W. III, and Hood, L. (1985). Structure of a gene encoding a murine thymus leukemia antigen, and organization of *Tla* genes in the BALB/c mouse. J. Exp. Med. **162**, 528-545.

Flaherty, L. (1981). *Tla*-region antigens. In *The Role of the Major Histocompatibility Complex in Immunobiology*, ed. M. Dorf, pp. 33-57. New York: Garland STPM.

Goodenow, R. S., McMillan, M., Orn, A., Nicolson, M., Davidson, N., Frelinger, J. A., and Hood, L. (1982a). Identification of a BALB/c $H\text{-}2L^d$ gene by DNA-mediated gene transfer. Science **215**, 677-679.

Goodenow, R. S., McMillan, M., Nicolson, M., Sher, B. T., Eakle, K., Davidson, N., and Hood, L. (1982b). Identification of the class I genes of the mouse major histocompatibility complex by DNA-mediated gene transfer. Nature **300**, 231-237.

Hammerling, U., Ronne, H., Widmark, E., Servenius, B., Denaro, M., Rask, L., and Peterson, P. A. (1985). Gene duplications in the TL region of the mouse major histocompatibility complex. EMBO J. **4**, 1431-1434.

Hood, L., Steinmetz, M., and Malissen, B. (1983). Genes of the major histocompatibility complex of the mouse. Ann. Rev. Immunol. **1**, 529-568.

Jaulin, C., Perrin, A., Abastado, J.-P., Dumas, B., Papamatheakis, J., and Kourilsky, P. (1985). Polymorphism in mouse and human class I *H-2* and *HLA* genes is not the result of random independent point mutations. Immunogenetics **22**, 453-470.

Klein, J. (1979). The major histocompatibility complex of the mouse. Science **203**, 516-521.

Klein, J., Figueroa, F., and David, C. S. (1983). *H-2* haplotypes, genes and antigens: Second listing. II. The *H-2* complex. Immunogenetics **17**, 553-596.

Kress, M., Cosman, D., Khoury, G., and Jay, G. (1983). Secretion of a transplantation related antigen. Cell **34**, 189-196.

Kvist, S., Roberts, L., and Dobberstein, B. (1983). Mouse histocompatibility genes: structure and organisation of a K^d gene. EMBO J. **2**, 245-254.

Lader, E., Clark, B. T., Jhanwar, S. C., Chaganti, R. S. K., and Bennett, D. (1985). Definitive chromosomal location of the H-2 complex by in situ hybridization to pachytene chromosomes. Immunogenetics **22**, 49-54.

Lalanne, J.-L., Transy, C., Guerin, S., Darche, S., Meulien, P. and Kourilsky, P. (1985). Expression of class I genes in the major histocompatibility complex: identification of eight distinct mRNAs in DBA/2 mouse liver. Cell **41**, 469-478.

Maloy, W. L., Coligan, J. E., Barra, Y., and Jay, G. (1984). Detection of a secreted form of the murine H-2 class I antigen with an antibody against its predicted carboxyl terminus. Proc. Natl. Acad. Sci. USA **81**, 1216-1220.

Mellor, A. L., Weiss, E. H., Kress, M., Jay, G., and Flavell, R. A. (1984). A nonpolymorphic class I gene in the major histocompatibility complex. Cell **36**, 139-134.

Mellor, A. L., Antoniou, J., and Robinson, P. J. (1985). Structure and expression of genes encoding murine Qa-2 class I antigens. Proc. Natl. Acad. Sci. USA **82**, 5920-5924.

Moore, K. W., Sher, B. T., Sun, Y. H., Eakle, K. A., Hood, L. (1982). DNA sequence of a gene encoding a BALB/c mouse L^d transplantation antigen. Science **215**, 679-682.

Ploegh, H. L., Orr, H. T., and Strominger, J. L. (1981). Major histocompatibility antigens: the human (HLA-A, -B, -C) and murine (H-2K, H-2D) class I molecules. Cell **24**, 287-299.

Pontarotti, P. A., Mashimo, H., Zeff, R. A., Fisher, D. A., Hood, L., Mellor, A., Flavell, R. A., and Nathenson, S. G. (1986). Conservation and diversity in the class I genes of the major histocompatibility complex: sequence analysis of a Tla^b gene and comparison with a Tla^c gene. Proc. Natl. Acad. Sci. U.S.A. in press.

Potter, M. (1985). The history of the BALB/c family. Current Topics in Microbiol. and Immunol. **122**, 1-5.

Reyes, A. A., and Wallace, R. B. (1982). A comparison of the coding and 3'-noncoding DNA sequences of several murine transplantation antigens. Immunogenetics **16**, 265-268.

Rogers, J. H. (1985). Family organization of the mouse *H-2* class I genes. Immunogenetics **21**, 342-353.

Sanger, F., Hicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**, 5463-5467.

Sher, B. T., Nairn, R., Coligan, J. E., Hood, L. E. (1985). DNA sequence of the mouse *H-2D*^{d} transplantation antigen gene. Proc. Natl. Acad. Sci. USA **82**, 1175-1179.

Steinmetz, M., Moore, K. W., Frelinger, J. G., Sher, B. T., Shen, F.-W., Boyse, E. A., and Hood, L. (1981). A pseudogene homologous to mouse transplantation antigens: transplantation antigens are encoded by eight exons that correlate with protein domains. Cell **25**, 683-692.

Steinmetz, M., Winoto, A., Minard, K., and Hood, L. (1982). Clusters of genes encoding mouse transplantation antigens. Cell **28**, 489-498.

Steinmetz, M., Malissen, M., Hood, L., Orn, A., Maki, R.A., Dastoornikoo, G. R., Stephan, D., Gibb, E., and Romaniuk, R. (1984). Tracts of high or low sequence divergence in the mouse major histocompatibility complex. EMBO J. **3**, 2995-3003.

Steinmetz, M., Stephan, D., and Lindahl, K. F. (1986). Gene organization and recombinational hot spots in the murine major histocompatibility complex. Cell, in press.

Strauss, E. C., Kobori, J. A., Siu, G., and Hood, L. E. (1986). Specific-primer-directed DNA sequencing. Anal. Biochem., in press.

Stroynowski, I., Orn, A., Goodenow, R. S., McMillan, M., Forman, J., Brayton, P. R., Frelinger, J., and Hood, L. (1984). Cytotoxic T lymphocytes recognize determinants on the BALB/c H-2L^{d} molecule controlled by α1 and α2 but not α3 external domains. Immunogenetics **20**, 141-154.

Stroynowski, I., Clark, S., Henderson, L. A., Hood, L., McMillan, M., and Forman, J. (1985). Interaction of α1 with α2 region in class I MHC proteins contributes determinants recognized by antibodies and cytotoxic T cells. J. Immunol. **135**, 2160-2166.

Sun, Y. H., Goodenow, R. S., and Hood, L. (1985). Molecular basis of the dm1 mutation in the major histocompatibility complex of the mouse: a *D/L* hybrid gene. J. Exp. Med. **162**, 1588-1602.

Tewarson, S., Zaleska-Rutczynska, Z., Figueroa, F., and Klein, J. (1983). Polymorphism of ***Qa*** and ***Tla*** loci of the mouse. Tissue Antigens **22**, 204-212.

Weiss, E., Golden, L., Zakut, R., Mellor, A., Fahrner, K., Kvist, S., and Flavell, R. A. (1983). The DNA sequence of the $H\text{-}2K^b$ gene: evidence for gene conversion as a mechanism for the generation of polymorphism in histocompatibility antigens. EMBO J. **2**, 453-462.

Weiss, E. H., Golden, L., Fahrner, K., Mellor, A. L., Devlin, J. J., Bullman, H., Tiddens, H., Bud, H., and Flavell, R. A. (1984). Organization and evolution of the class I gene family in the major histocompatibility complex of the C57BL/10 mouse. Nature **310**, 650-655.

Winoto, A., Steinmetz, M., and Hood, L. (1983). Genetic mapping in the major histocompatibility complex by restriction enzyme site polymorphisms: Most mouse class I genes map to the Tla complex. Proc. Natl. Acad. Sci. USA **80**, 3425-3429.

Yokoyama, K., and Nathenson, S. G. (1983). Intramolecular organization of class I H-2 MHC antigens; localization of the alloantigenic determinants and the β_2m binding site to different regions of the H-2 K^b glycoprotein. J. Immunol. **130**, 1419-1425.

Zinkernagel, R. M., and Doherty, P. C. (1979). MHC restricted cytotoxic T cells: Studies on the biological role of polymorphic major transplantation antigens determining T cell restriction specificity, function and responsiveness. Adv. Immunol. **27**, 51-75.

Zuniga, M. C., Malissen, B., McMillan, M., Brayton, P. R., Clark, S. S., Forman, J., and Hood, L. (1983). Expression and function of transplantation antigens with altered or deleted cytoplasmic domains. Cell **34**, 535-544.

Zuniga, M. C., and Hood, L. E. (1986). Clonal variation in cell surface display of an H-2 protein lacking a cytoplasmic tail. J. Cell Biol. **102**, 1-10.

Figure 1. Map of the class I genes in the BALB/c MHC.

34 class I genes represented by a total of at least 120 cosmid and phage clones have been linked into six clusters and mapped into the *K*, *D*, *Qa* and *Tla* regions (Steinmetz et al., 1982; Steinmetz et al., 1984; Rogers, 1985; Fisher et al., 1985; Chapter Two). Each gene is denoted by a solid box which is not drawn to scale. The chromosomal location of the "37" gene is unknown and is tentatively placed in the Tla region due to its homology to genes in the Tla region. The I region contains the class II gene family and the S region contains genes coding for certain complement components (C2, Bf, Slp, C4) and a steroid biosynthesis enzyme (21-hydroxylase: 21-OH), and will not be discussed further.

The large number of BALB/c class I clones have been isolated from several recombinant libraries made of DNA from several BALB/c substrains: a cosmid library (Steinmetz et al., 1982) made of BALB/cCum and BALB/crgl DNA; a lambda library (Steinmetz et al., 1982) made of BALB/crgl; a cosmid library (Sun et al., 1985) made of BALB/cJ (purchased from the Jackson Laboratory, Bar Harbor, Maine) DNA; a cosmid library (Steinmetz et al., 1984) and a lambda library (Chapter Two) made of DNA from a colony of BALB/cJ maintained at the Basel Institute for a number of years. In view of the reported divergence between the BALB/c substrains (Mellor et al., 1985; see Potter, 1985), it is of concern whether the large number of class I genes is an artifact of the substrain differences. However, they do appear to be truly present in the BALB/c genome for the following reasons. First, the block of DNA containing T11-T17 is absent from the C57BL/10 mouse, but Southern blot analysis of genomc DNA demonstrated that these genes are represented in the BALB/c genome and have the correct restriction fragment length as predicted from the cloned DNA (Fisher et al., 1985). Second, the differences between the substrains are likely to be subtle, whereas the different cloned genes, although similar between the proposed

duplicated regions, have diverged to a considerable degree. The most striking homology is between the D and L genes. Both of them, however, are known to exist in the BALB/c genome.

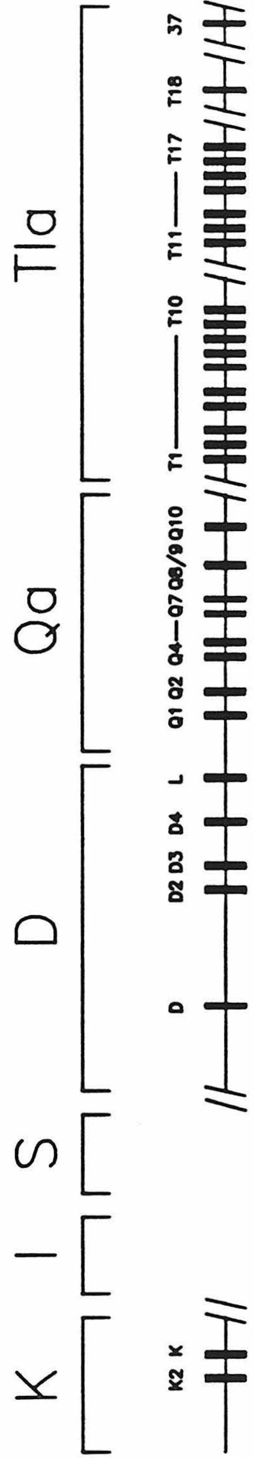

K
I
S
D
Qa
Tla
K2 K
D
D2 D3 D4 L
Q1 Q2 Q4—Q7 Q8/9 Q10
T1—T10 T11—T17 T18 37

Figure 2. Strategy for the sequencing of class I transmembrane exons.

An 18 nucleotide sequence, located approximately 70 nucleotides 5' to the end of the fourth exon, was identified to be invariant among all of the sequenced class I genes. A synthetic oligonucleotide derived from this sequence was used as primer to sequence through the end of the fourth exon, across the short fourth intron, and into the transmembrane exon in a single sequencing reaction. The exons of a typical class I gene are shown in filled boxes. The location of the sequencing primer is shown as a dot above the fourth exon, and the direction of the sequencing reaction is indicated by the waved line and arrow.

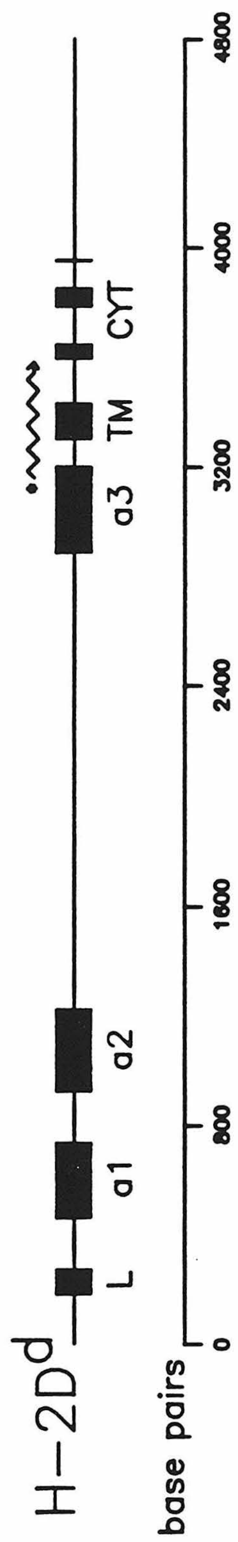

H-2Dd
L
a1
a2
a3
TM
CYT
base pairs
0
800
1600
2400
3200
4000
4800

Figure 3. Sequence comparison of 30 transmembrane sequences.

The first line, labeled TM, is a consensus sequence determined from the first 19 sequences, which constitute the large subfamily. All sequences were aligned and a minimum number of gaps inserted to achieve maximum homology. Homology to the consensus sequence is indicated by a dot. A solid line separates the subfamilies. The shaded area indicates the transmembrane exon. The fourth exon ends at position 54 and is indicated by a large arrow. The homologies between the different subfamilies drops sharply within the exon, and analyses using the Best-Fit and Dot-Matrix computer programs do not detect any other alignment that will give significantly higher homologies.

Nonrandom base substitutions are denoted by small arrows above the consensus sequence. These were determined by the following criterion: the number of times a particular base change occurring independently is divided by the number of total changes at that position, and scores above 1/3, which is the expected value for random changes, were registered as nonrandom changes. The same changes occurring in the duplicated pair are counted as a single occurrence. The positions where a CpG dinucleotide occurs is denoted by **x**. These analyses included only sequences of the large subfamily (group 1).

The sequences corresponding to the oligonucleotides are underlined.

```
TM    1  ACACATGCCATGTGTACCATGAGGGGCTGCCTGAGCCCCTCACCCTGAGATGGGGTAAGGAGGGTGTGGGTGCAGAGCTGGGG-TCAGGGAAAGCTGGAGC--CTTCTGCA--GACCCTG  120

TM  121  A-GCTGGTCAGGGCTGAGAGCTGGGGTCATGACCCTCACCT--------------------------TCATTTCCTGTACCTGTCCTTCCCAGAGCCTCCTCCATCCACTGTCTCCAACAT  240

TM  241  GGTAATCATTGCTGTTCTGGTTGTCCTTGGAGCTGTGGTCATCATTGGAGCTGTGGT------------------GGCTTTTGTGATGAAGAGGAGGAGAAACACAGGTAGGAAAGGGCA  360

TM  361  GGGTCTGAGTTTCTCTCAGCCT-CCTTTAGAAGTGTGCTCTGCTCATTAATGGGGAACACAGCCACACCCCACATTG-CTACTGTCTCTAACTGGGTCTGCTGTCAGTTCTGGGAACTT  480
```

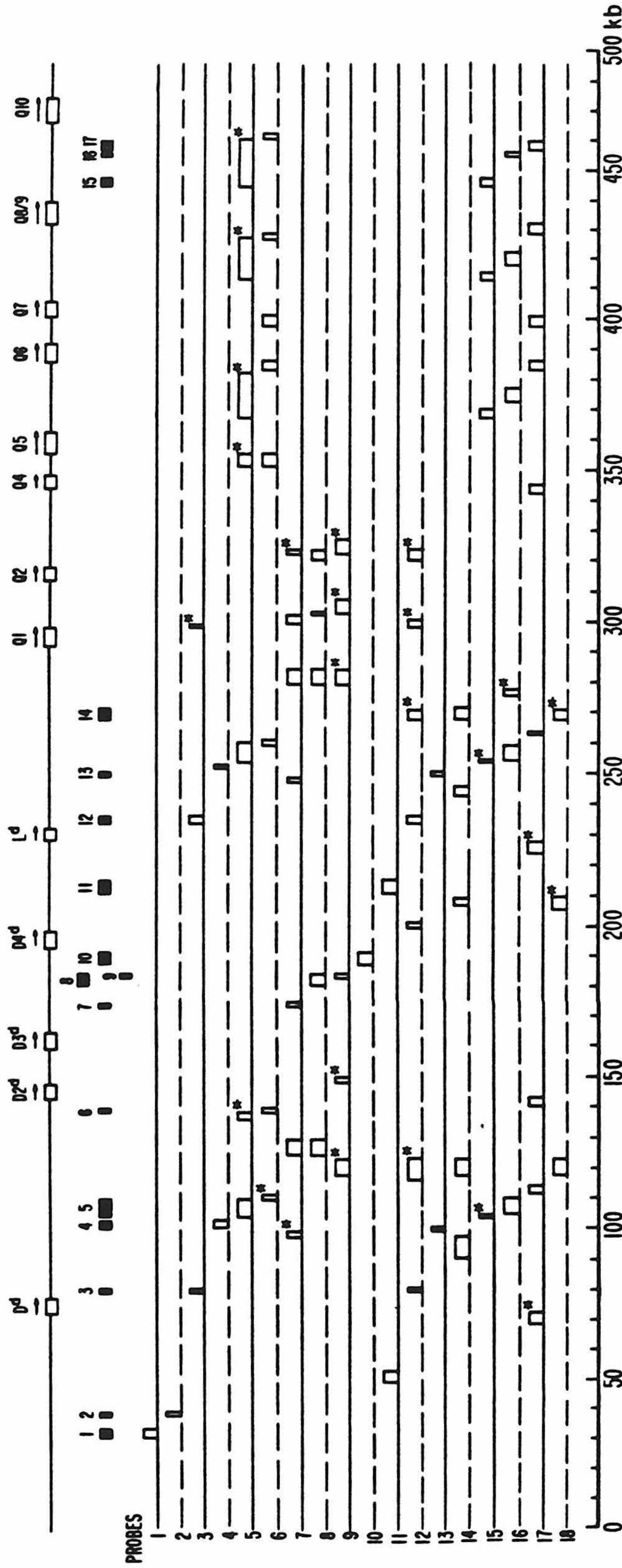
Figure 4. Translation of the transmembrane exons.

Analyses of complete sequences of class I genes have shown that, as a rule, the exons always start with the second nucleotide of a codon (see Hood et al., 1983). Since the last nucleotide of the previous exon were not known for all of the genes, the translations start with the third nucleotide. Charged amino acid residues are shaded. Homology to the K sequence is indicated by a dot at that position.

```
K      LPPSTVSNTVIIAVLVVLGAAIVTGAVV------AFVMKMRRNT
K2     P.L..D..M.T.........TV..IV...------.....R....
D      P.S..KT.......P.....VVIL..M------.....R....
Q5     P.S..KT.......P.....VVIL...------.....R....
D2     P..Y....M...........LAIF...------.....R..H.
D4     P....DT.MA...F......V..I..M.------...L.R....
L      P....D.YM..V...G....MAII...------.....R....
Q1     P..Y....M...........V..I...IIGVMVS.....R...K
Q2     P.......M...........V.IIA..------.....RG...
Q4     P.......MANV........WPSLQLWW------LLX
Q6     P.......MANV.I....V.WPSLQLWW------LLX
Q7     P..Y....MAT...V.D...VAII...------....NR.X
Q8/9   P..Y....MAT..IV.D...VAII...------....NR.X
Q10    P....D.IMSH..D.----LWPSLKLWW------YLX
T4     P...MD..M..V........V.II..M.------...L.S.KKI
T6     P....D..M..M...FFFFFFFFFC
T8     PL..AD..M....A...F..V.II...------..M..R.N-.
T16    P.S...FPM.........P..VVINV...------...--R.N-K
"37"   P.......M...........V.IL...------.....R..HI
T10    P.YY..FKMAT...VDD..QVAII.P.A--------T..NR.SX
T1     P.Q.SMPTRT.VG.VLGAMVI-GLKSRNV--------M.W..K.K
T3     P.QTSMP.RTTVRA.LGAMIILGFMSGS--------VM.W..K.N
T13    P.Q.SMP.RTTVRA.LGAMIILGFMSGS--------VM.W..K.N
T2     P.QPIIFIRT.VGA--....GVIL.FII---GGVKMW..KK.K
T12    P.X
T5     P.Q.IILIRT.VGA--....GKDVSGSN--QAWX
T7     P....-LH.SNF.LS...L.VTFIYLWW-----FFCLFVCFWLV
T15    P....-LHRSNF.LS...L.VTFNLPM.-----VFCLFVCFWLV
T9     PAW----YQKPWIWI.AMVFILLIICLC------VVCSC.KK.A
T17    PAW----YQKPWIWI.AMVFILFIICLC------VVCSC.KK.A
                 |          |          |          |
```

Figure 5. Hybridization of the 23 flanking region probes to the class I region.

A molecular map of the five cluster of class I genes are shown (for restriction maps, see Steinmetz et al., 1982; Steinmetz et al., 1984; Fisher et al., 1985; Chapter two). The entire class I region is represented by 33 cosmid clones spaning a total of 1000 kb (see Materials and Methods). The class I genes are shown in open boxes with arrows above to indicate transcription orientation. The position from where the various probes were isolated are shown in solid boxes just below the map, and are consecutively numbered 1 to 23. The positions of the smallest fragment that hybridized to each probe are shown in open boxes on alternating solid and dashed lines labeled with the probe number. A * indicates weak hybridization. The map is drawn to scale. A portion of the hybridization data has been previously reported (Chapter Two).


PROBES
0
50
100
150
200
250
300
350
400
450
500 kb

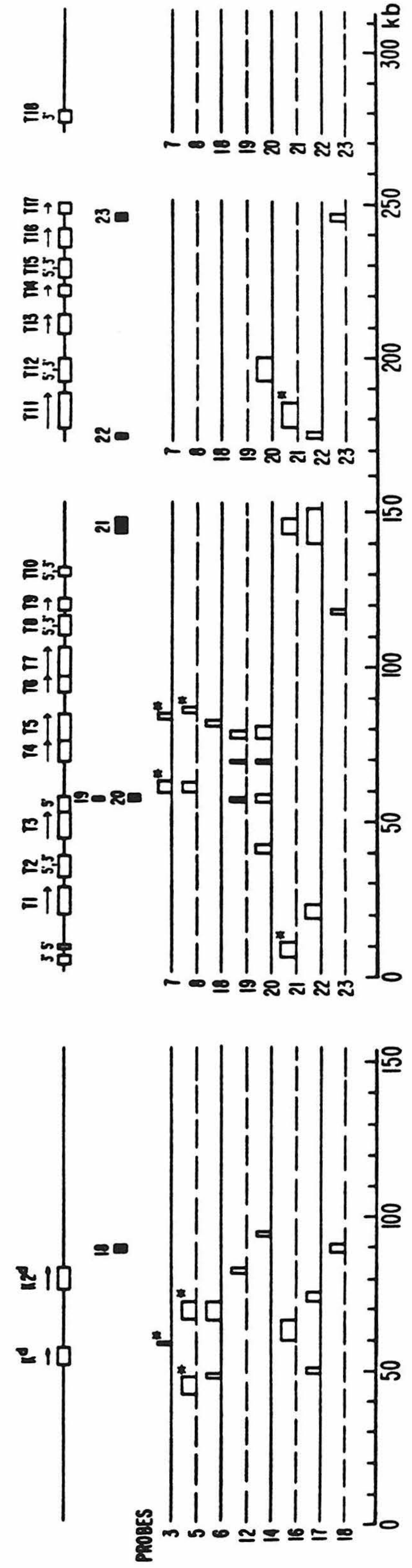

PROBES
0
50
100
150
200
250
300 kb

Table 1. Pair-wise homology comparison of the transmembrane exons.

The comparison includes the exon only (nucleotide 233-367, see Figure 3). Each gap is counted as one mismatch irrespective of its length, since it is assumed to have occurred in a single event. The dark solid lines separate the different subfamilies. The shaded area indicates the comparisons between members of the first group. Hatched blocks indicate the duplicated pair of genes.

	K	K2	D	D2	D4	L	Q1	Q2	Q4	Q5	Q6	Q7	Q8/9	Q10	T4	T6	T8	T16	"37"	T10	T1	T3	T13	T2	T12	T5	T7	T15	T9
K2	86																												
D	82	84																											
D2	84	84	84																										
D4	87	90	85	84																									
L	82	85	85	87	85																								
Q1	88	88	83	89	89	84																							
Q2	88	89	85	89	89	89	90																						
Q4	82	85	81	89	85	91	85	91																					
Q5	83	84	98	85	85	85	84	86	82																				
Q6	81	84	80	88	85	90	84	88	97	81																			
Q7	83	86	82	89	86	87	86	89	90	83	89																		
Q8/9	82	85	81	88	86	86	85	88	89	82	90	99																	
Q10	71	76	74	77	77	82	73	80	81	75	80	79	78																
T4	82	85	82	83	87	85	86	88	85	83	84	82	81	74															
T6	89	89	78	87	87	85	89	93	85	78	83	83	83	76	93														
T8	85	86	85	84	85	85	84	88	82	85	81	82	82	76	85	87													
T16	82	82	80	82	81	81	82	86	81	81	79	81	81	71	81	85	82												
"37"	88	88	87	91	89	87	92	93	91	88	90	88	87	77	91	93	89	86											
T10	62	67	61	69	64	67	66	67	68	62	67	78	77	59	61	65	62	64	67										
T1	52	48	47	45	51	47	50	47	47	46	47	47	46	43	47	61	45	47	48	40									
T3	51	52	46	46	49	46	50	46	47	46	47	46	46	44	45	57	47	46	47	41	87								
T13	52	53	47	47	50	47	51	47	48	46	48	47	46	45	46	59	48	47	48	42	88	99							
T2	54	56	60	56	57	57	55	56	56	58	55	56	56	52	58	52	54	52	58	49	49	46	46						
T12	52	55	60	55	56	56	53	55	54	58	55	55	56	50	56	48	52	52	56	49	47	44	44	98					
T5	61	60	61	56	63	60	58	63	63	60	59	60	56	55	63	52	58	60	61	51	67	61	63	95	91				
T7	55	56	54	53	58	56	55	58	58	55	59	59	59	52	58	65	54	56	57	49	31	32	33	42	41	57			
T15	54	55	53	52	58	55	56	57	58	54	58	59	59	51	57	65	53	57	56	49	32	33	34	43	42	59	96		
T9	37	40	34	34	37	38	35	36	35	34	36	38	38	33	34	29	35	33	35	32	35	40	40	34	34	18	25	26	
T17	37	41	34	34	37	38	35	36	36	34	37	38	38	34	34	29	35	33	35	32	34	40	40	33	33	18	27	27	99

CHAPTER FOUR

A RAPID METHOD FOR RESTRICTION SITE MAPPING OF COSMIDS AND PLASMIDS

A RAPID METHOD FOR RESTRICTION SITE MAPPING OF COSMIDS AND PLASMIDS

By Y. Henry Sun and Leroy Hood

From the Division of Biology, California Institute of Technology
Pasadena, California 91125

INTRODUCTION

Restriction site mapping is one of the most fundamental and important ways of characterizing a cloned DNA fragment. A number of strategies have been employed (1-6), but none of them offer a simple and satisfactory method for restriction site mapping. We have developed a restriction site mapping method which is rapid, simple, circumvents the difficulties of existing methods, and has general applicability. The method is a modification of those by Smith and Birnstiel (1) and by Rackwitz *et al.* (5).

Our method was originally developed for mapping large DNA fragments (30-45 kb) cloned in cosmid vectors. The principle can be easily applied to a wide variety of recombinant clones. To illustrate the principle and procedures of the method, a cosmid clone in the vector pTL5 was used as an example.

MATERIALS AND METHODS

ENZYMES AND CHEMICALS. Restriction endonucleases were purchased from New England Biolabs (NEB) or Boehringer-Mannheim Biochemicals (BM). T4 polynucleotide kinase was from NEB. γ-P^{32}-ATP (7000 Ci/mmole) was purchased from New England Nuclear Products (NEN).

CLONES. The cosmid vector pTL5 has been described (7,8). The cosmid clone dm1-7.2 was isolated from a library constructed of B10.D2-H-2^{dm1} mouse liver DNA cloned into the BglII site of pTL5 (9).

OLIGONUCLEOTIDES. Two 40 mer oligonucleotides, S_A (5'-CTGGGCTG-CTTCCTAATGCAGGAGTCGCATAAGGGAGAGC-3') and S_B (5'-CCACCGGAAG-GAGCTGACTGGGTTGAAGGCTCTCAAGGGC-3'), were synthesized in the Caltech Microchemical Facility with an Applied Biosystems 380A synthesizer and purified according to the procedure of Strauss *et al.* (10).

LINEARIZATION. The SalI restriction site is used here as an example.

Other sites, as discussed in the text, can also be used. About 100 ng of the cosmid clone is first digested with SalI to determine whether the insert contains more than one site. About 10 μg of cosmid DNA is digested with SalI to completion, extracted once with buffered phenol, twice with chloroform, precipitated with ammonium-acetate and ethanol, and resuspended in TE to a final concentration of 100 μg/ml.

PARTIAL DIGESTION. We have found that varying the enzyme concentration is easier to control than varying the reaction time. We have also found the conditions for partial digestion difficult to reproduce exactly. The following strategy avoids the need of reproducing the partial digestions. We do a series of 5-6 digestions with twofold serial dilution of each enzyme. Restriction enzymes were serially diluted in 1X digestion buffers to the needed concentrations. Typically. 200 ng of SalI-digested DNA was digested in each reaction of 20 μl at 37°C for 50 minutes. The reactions were then stopped by adding 5 μl of stop buffer (0.1 M EDTA, 0.1% SDS, 15% Ficoll, 0.15% bromophenol blue, 0.15% xylene cyanol). Many reactions can be started and stopped synchronously by pipeting the drop of enzyme solution or stop buffer onto the inside of the cap on the Eppendorf tube, closing the cap, and spinning all tubes in the microfuge briefly to bring down the drop into contact with the sample.

A third of each reaction (70 ng) was run on a 0.6% agarose gel (17 cm x 19 cm, with two rows of 25 wells, so many samples can be checked with a minimum number of gels) in TPE buffer (80 mM Tris-phosphate, pH 8.0, 8 mM EDTA). The remaining portions were temporarily stored at -20°C. The reactions that gave the best range of partial digests were pooled. Since each of the partial digests is sufficient for at least three runs, the digestions do not need to be repeated.

MAPPING GEL. 70-100 ng of each pooled sample was run on a 0.6 % horizontal agarose gel (17 x 19 cm, 20 wells) in TPE buffer at 1 V/cm. Good

resolution depends on low voltage gradient (1 V/cm) and loading of no more than 100 ng per lane. Usually two loadings that run different lengths of time were performed to get good resolution from a few kb to 35 kb. Thin, vertical agarose gels casted between sandblasted glass plates have been tried. They gave better resolution but are much more difficult to prepare.

Lambda DNA digested separately with the restriction enzymes HindIII, XhoI, SmaI, EcoRI, and KpnI and uncut lambda DNA were used as size markers. The marker DNA was heated at 68°C for 5 min immediately before loading onto the gel. After electrophoresis, the gel was stained with ethidium bromide (0.5 μg/ml), photographed under UV-illumination, and blotted according to the procedure of Southern (11) except 1 M ammonium acetate was used as neutralizing and transfer buffer (12). Two filters can be blotted from each gel, either sequentially or by the sandwich method (13).

HYBRIDIZATION. Each filter was hybridized to one labeled oligonucleotide. The oligonucleotides have a calculated Tm of 92°C (S_A) and 94°C (S_B) at 0.9 M Na^+ (5X SSPE). They are end-labeled with γ-P^{32}-ATP and polynucleotide kinase (14), and purified over a Sephadex G-50 spun-column (14). Hybridization was at 5X SSPE(1X = 0.18 M NaCl, 1 mM EDTA, 10 mM NaH_2PO_4, pH 7.4), 5X Denhardt's solution (1X = 0.02% Ficoll, 0.02% bovine serum albumin, 0.02% polyvinyl pyrolidone), 10% dextran sulfate, 10^6 cpm/ml, 60°C for 12-16 hr. The filters were washed at 3X SSPE at 60°C and autoradiographed with Kodak XAR-5 film at -70°C with intensifying screen for 12 hr.

RESULTS

The original Smith and Birnstiel mapping method (1) requires radioactively labelling of both ends of a restriction fragment, cleavage at an unique internal restriction site that is asymmetrically located, gel purification of the resulting

two fragments, each now labeled at only one end, and partial digestion of each of the fragments with a second restriction enzyme. To simplify this laborious process, we originally tried to take advantage of two closely spaced restriction sites, SalI and ClaI, in the cosmid vector pTL5 (Fig. 1). The two recognition sites occur very rarely in eukaryotic DNA. Many of the cosmid clones that we isolated contain inserts that have neither recognition sites. Therefore, the cosmid can first be linearized with SalI, end-labeled, and digested with ClaI. This will generate a small fragment of 627 bp derived from the cosmid vector and a large fragment containing the remainder of the vector and the entire insert (Fig. 1). Each fragment is labeled at only one end. Similarly, the clone can be first cleaved with ClaI, end-labeled and then cleaved with SalI, thus generating the same large fragment but labeled at the other end. Without gel purification, the digest is subjected to partial digestion with the enzymes used for mapping. Because the small fragment is much shorter than the vector arm on the larger fragment, it will not interfere with the analysis. This eliminated the need for gel purification. However, all of the partial fragments give a faint signal, presumably due to internal labeling at nicks generated during DNA preparation. Although mapping of cosmid-sized clones are difficult due to this background, this mapping strategy can still be useful for plasmid or M13 phage clones.

To get around the background problem, end-specific oligonucleotides were used to hybridize to the unlabeled sample DNA, thereby providing end-specific labeling. Rackwitz *et al.* (5) have used two 12 mer oligonucleotides, which are complementary to the left and right lambda cos sequence, respectively, to map lambda clones. Cosmid clones can be similarly mapped if they were linearized at the cos site by the lambda terminase. Since the lambda terminase is not readily available, restriction sites that occur in the cosmid vector but not in the insert, such as SalI and ClaI, were used to linearize the cosmid DNA.

The procedure is outlined in Fig. 2. The clone was first linearized with a restriction enzyme that cleaves only once in the clone, e.g., SalI. The linear DNA was then subjected to partial digestion with each of the restriction enzymes whose sites are to be mapped. An aliquot of each reaction was checked by gel electrophoresis. The remaining portion of the ones that gave the best range of partial digest were pooled and run on a 0.6% agarose gel. The gel was blotted onto nitrocellulose and hybridized to one of the two end-labeled oligonucleotides.

The two 40 mer oligonucleotides are complementary to the sequences on either side of the SalI site in the vector. Therefore, each of them will hybridize to only one end of the linearized cosmid and therefore to a series of partial fragments, all starting from that same terminus but ending at different lengths. Each end point represents a restriction site. Upon autoradiography, a banding pattern much like the DNA sequencing ladder will show up, and the position of the restriction sites relative to the terminus and to each other can be read off directly.

An example is shown in Fig. 3. The resolution was good up to 35 kb. With the two oligonucleotides each reading 35 kb from one end and overlap at the center of the clone, the entire clone was easily mapped. More accurate size of each fragment can be obtained from measurements of the complete digests. The map generated by this method is in excellent agreement with that generated by the conventional double-digestion method.

DISCUSSION

All existing mapping methods have some disadvantages. The double digestion method (2) is laborious in that many different combinations of double digestions have to be performed, the trial-and-error analysis process is very time consuming, is particularlly difficult when mapping the many sites of a frequent-

cutter enzyme, and very precise size measurement is usually required since part of the logic is to add up the sizes of double-digest fragments to account for single-digest fragments. The sequential double-digestion using gel-purified fragments (3) is laborious, although it made the logic of piecing fragments together much easier. The southern cross method (6) has difficulty when there are repetitive elements within the clone. The partial digestion method by Smith and Birnstiel (1) simplifies greatly the data analysis step, but requires a unique second restriction site within the fragment to be mapped, which cannot always be fulfilled, and also requires laborious gel-purification of end-labeled fragments. Internal labeling at nicks in the DNA molecules generates a background that often obscures the signal, especially when mapping larger DNA fragments. Preferential labeling of one end (4) is applicable only when the two sticky ends start with different nucleotides. The indirect-labeling method (5) circumvents the above problems, but is limited to lambda clones and, only when lambda terminase becomes available, can be extended to cosmid clones. Our method extends the applicability to virtually any type of cloning vectors.

The requirement for our method is the presence of a restriction site in the vector, but not more than once in the insert. SalI and ClaI were chosen for the cosmid vector pTL5. By calculating dinucleotide frequency (15), these two sites should occur in the mouse genome about every 31 kb and 18 kb, respectively. In fact, when mouse genomic DNA was digested with either of these two enzymes, the majority of the fragments were larger than 50 kb (data not shown). In our experience with mouse cosmid clones from the MHC class I genes, SalI appeared only 20 times and ClaI appeared only 14 times in 1000 kb of DNA. The probability of both sites occurring more than once in the same clone is very low. Therefore, four oligonucleotides, two on each side of the SalI site and two on each side of the ClaI site, will be sufficient to map most cosmid clones in the vector pTL5. In

fact, two of the four oligonucleotides, S_B and C_A, are sufficient for clones linearized with either SalI or ClaI (see Fig. 2), with the following caution. BamHI, HindIII and EcoRI sites lie between the SalI and ClaI sites; therefore, the BamHI-, HindIII- or EcoRI-digested fragments will be slightly shorter than those from other enzyme digests.

Since the SalI and ClaI sites in pTL5 are associated with the tet^r gene that was derived from the popular plasmid vector pBR322, the two sets of oligonucleotides can also be used in any pBR322-derived tet^r cloning vectors. Other rare occurring restriction sites can be chosen for other vectors, and a limited number of oligonucleotides corresponding to these sites will cover most of the commonly used cloning vectors. For example, for the popularly used vectors M13mp18, mp19, pUC18, and pUC19, the clones can be cut with one of several restriction enzymes, whose recognition sites lie upstream or downstream of the polylinker and which does not cut in the insert, and the commercially available sequencing or reverse sequencing primers can then be used for the mapping.

This method is easy and rapid. Only the two oligonucleotides need to be radioactively labeled instead of labeling every fragment. Since the method depends on specific hybridization, relatively crude plasmid preparations can also be used to save more time and work. With the protocol presented here, many samples can be processed at the same time. We have easily mapped 10 cosmid clones with BamHI, HpaI, and KpnI in less than two weeks. It is particularly useful when mapping with enzymes that cut frequently.

REFERENCES

1. Smith, H.O. and Birnstiel, M.L. (1976) Nucleic Acids Res. **3**, 2387-2400.
2. Danna, K.J., Sack, G.H., Jr., and Nathans, D. (1973) J. Mol. Biol. **78**, 363-376.
3. Parker, R.C. and Seed, B. (1980) *Methods in Enzymology*, vol. 65, Grossman, L. and Moldave, K., eds. (Academic Press, Inc., New York, New York), pp. 358-363.
4. Clark, A.J., Clissold, P.M., and Bishop, J.O. (1982) Gene **18**, 221-230.
5. Rackwitz, H.-R., Zehetner, G., Frischauf, A.-M., and Lehrach, H. (1984) Gene **30**, 195-200.
6. Potter, H. and Dressler, D., personal communication.
7. Lund, T., Grosveld, F.G., and Flavell, R.A. (1982) Proc. Natl. Acad. Sci. USA **79**, 520-524.
8. Steinmetz, M., Winoto, A., Minard, K., and Hood, L. (1982) Cell **28**, 489-498.
9. Sun, Y.H., Goodenow, R.S., and Hood, L. (1985) J. Exp. Med. **162**, 1588-1602.
10. Strauss, E.C., Kobori, J.A., Siu, G., and Hood, L.E. (1985) Anal. Biochem., in press.
11. Southern, E.M. (1975) J. Mol. Biol. **98**, 503-517.
12. Kafatos, F.C., Jones, C.W. and Efstratiadis, A. (1979) Nucleic Acids Res. **7**, 1541-1552.
13. Smith, G.E. and Summers, M.D. (1980) Anal. Biochem. **109**, 123-129.
14. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular Cloning. A Laboratory Manual, Cold Spring Harbor Laboratory, New York.
15. Swartz, M.N., Trautner, T.A., and Kornberg, A. (1962) J. Biol. Chem. **237**, 1961-1967.

Figure 1. A cosmid clone consisting of an insert about 30-45 kb in length cloned into the BglII site of the vector pTL5. (a) The clone is first linearized with SalI and end-labeled by T4 kinase. The open box denotes the insert, not shown to scale. The linear molecule is then cut with ClaI to generate two fragments each labeled at only one end. The large fragment is then subjected to partial digestions without purifying it away from the small fragment. (b) Same as in (a), except the order of SalI and ClaI digestion is reversed. This generates the same large fragment but labeled at the opposite end.

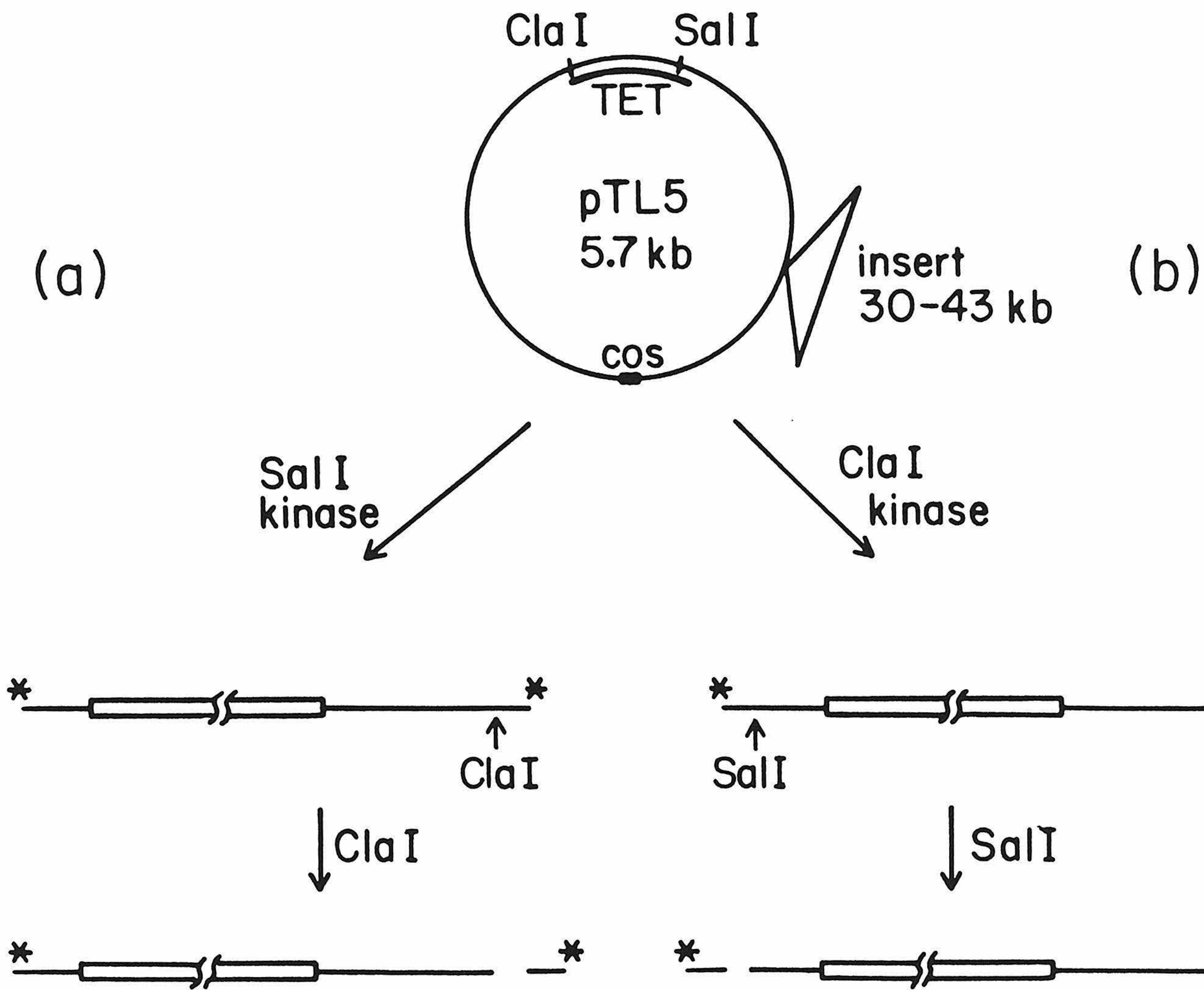
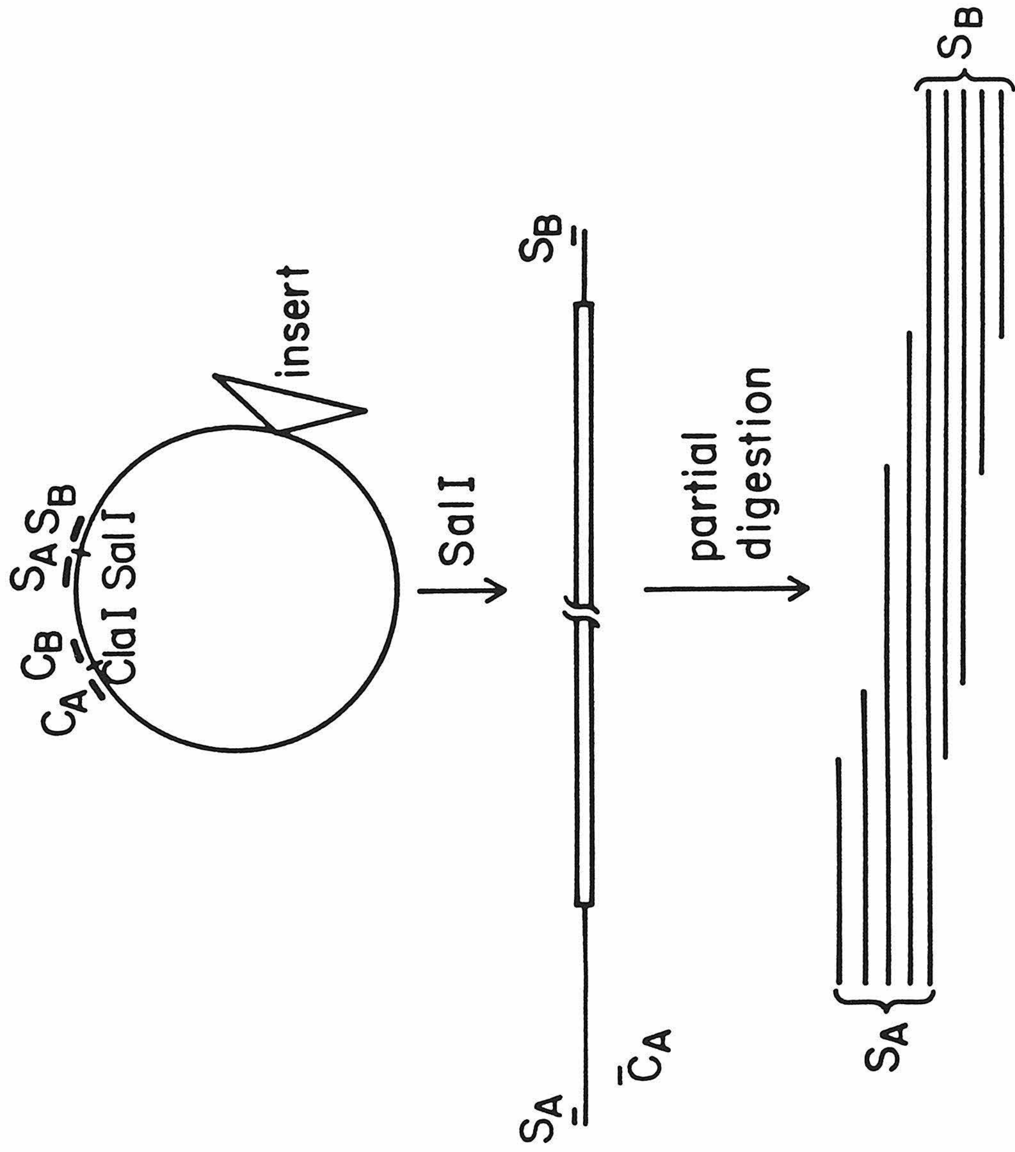
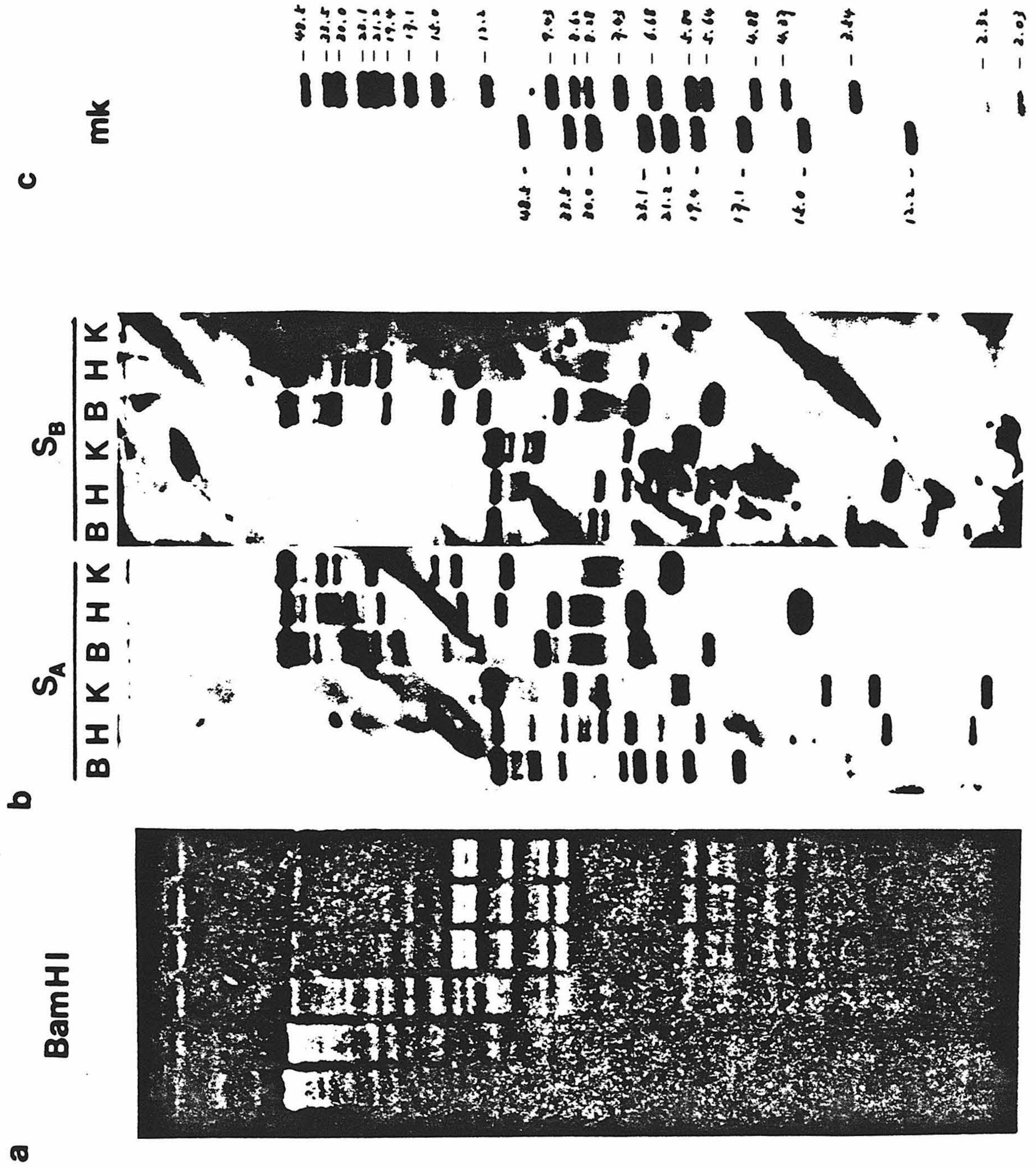

ClaI
SalI
TET
pTL5
5.7 kb
cos
insert
30-43 kb
(a)
(b)
SalI
kinase
ClaI
kinase
ClaI
SalI
ClaI
SalI

Figure 2. The position of the four oligonucleotides, S_A, S_B, C_A, and C_B, are shown (not to scale). The cosmid clone was first linearized with SalI and then subjected to partial digestions. The digests were separated by agarose gel electrophoresis and blotted onto nitrocellulose paper. Oligonucleotide S_A will hybridize to only those fragments containing the left end of the SalI site. Oligonucleotide S_B hybridizes only to those fragments containing the right end of the SalI site.



CA
CB
ClaI
SA SB
SalI
insert
SalI
SB
SA
CA
partial
digestion
SA
SB

Figure 3. (a) SalI linearized cosmid clone dm1-7.2 was subjected to partial digestion with BamHI, HpaI, and KpnI. Only the BamHI digests are shown: lanes 1 to 6 are, respectivley, 1/16, 1/8, 1/4, 1/2, 1, and 2 units of BamHI in each reaction of 200 ng DNA in 20 μl. (b) The appropriate partial digests of BamHI [e.g., reactions 2 and 3 in (a)], HpaI, and KpnI were pooled and 100 ng of each pooled digest were run on the mapping gel. The gel was blotted and hybridized separately with end-labeled S_A and S_B. Lanes 1-3 and 7-9 were run at 1 V/cm for 60 hr; lanes 4-6 and 10-12 were run at 1V/cm for 24 hr. (c) Size markers were also run for 60 and 24 hr, respectively. After blotting, these lanes were hybridized separately with labeled λ DNA.


a
BamHI
b
S_A
B H K B H K
S_B
B H K B H K
c
mk

CONCLUSION

The study of the major histocompatibility complex has always been a field of confusion, filled with complicated, ambiguous and controversial results and interpretations. Only in recent years, particularly with the application of molecular biology approaches and the rapid progress in the study of the T cell receptor, the many bits of information are beginning to fit into place to form a understanding of the true role of MHC in the immune system. As Jan Klein said: "It is as if molecular genetics has raised the curtain on the final act of the MHC play, and as in every good play this act is turning out to be short and packed with action. It truly was a good play: dead-earnest most of the time, hilarious on occasion, and thrilling all the time." (Klein, 1983).

Gene Organization

One of the most interesting and unexpected findings from the application of molecular biology approaches to the system is that there are a large number of sequences homologous to the H-2 class I genes. These genes have been isolated from recombinant libraries, and can be ordered into clusters of genes, based on overlapping restriction maps of the clones (Steinmetz et al., 1982). By examining for restriction fragment length polymorphism in congenic and recombinant congenic mouse strains, they were mapped into the various genetically defined regions of the MHC (Winoto et al., 1983). The cloned genes were transfected into mouse cell lines and assayed with a large panel of antisera and monoclonal antibodies to identify which known antigen each gene encodes (Goodenow et al., 1982). Genes of all well-defined class I antigens have been identified by this approach. DNA sequencing of most of these has confirmed these assignments by comparison with the protein sequences available for each of these antigens.

Thirteen clusters containing a total of 36 class I genes from the BALB/c mouse were originally reported (Steinmetz et al., 1982). Some of these clusters were later reported to contain sequences at their ends that are not contiguous

with the cluster in the genome, and probably are cloning artifacts. Eliminating these sequences allowed further overlapping of the clusters and genes (Rogers, 1985; Fisher et al., 1985), forming 9 clusters with 33 genes. To link up these clusters, the approach of chromosome walking was used. The approach is to isolate a low-copy number sequence from the end of a cluster and use it as a hybridization probe to screen a recombinant library. The clones isolated would overlap with the original clone and some would be an extension of the region already cloned. Similar studies were also done with the class II and class III genes. The new molecular map of the MHC contains about 1600 kb of cloned DNA with a total of 46 genes (see Figure 1 in Chapter Two). A 600 kb gene cluster spans the *H-2K* and *H-2I* regions and contains two class I (*K2*, *K*) and seven class II ($A_{\beta 3}$, $A_{\beta 2}$, A_{β}, A_{α}, $E_{\beta 2}$, E_{β}, E_{α}) genes (Steinmetz et al., 1986). Following a gap of unknown size, there is a 250 kb cluster of the *H-2S* region which contains four complement-related (*C2*, *Bf*, *Slp*, *C4*) and two 21-hydroxylase (*21-OH*) genes (Chaplin et al., 1983; Amor et al., 1985; White et al., 1984). This is followed at an unknown distance by a 500 kb cluster spanning the *H-2D* and *Qa* regions and containing five *D*-region (*D*, *D2-D4*, L) and eight *Qa*-region (*Q1*, *Q2*, *Q4-Q7*, *Q8/9*, *Q10*) genes (Chapter Two). Three gene clusters with 18 class I genes (*T1-T18*) (Rogers, 1985; Fisher et al., 1985) were mapped into the *Tla* region. Finally, the "37" gene (Lalanne et al., 1985) is homologous to a *Tla* region gene, but its genetic location is not known.

Gene Expressions

The finding of a large number of class I genes is unexpected. It is not known whether these genes are expressed, and if some of them are, what functions they play. Study of their expression has been difficult because these genes are highly homologous to each other, and therefore it is not easy to distinguish one from the others. With the set of gene-specific oligonucleotide recently developed (Chapter

Three), it is now possible to study the expression of individual genes at the level of transcription. In fact, preliminary studies have already identified the expression of several genes in several tissues (Hunt, S.W., personal communication).

Many antigens have been genetically mapped into the MHC. Some of these (H-2K, H-2D, H-2L, Qa-2, Qa-1, Tla-1,2) are well characterized serologically and biochemically, and all but one of the genes encoding them have been identified. Many others have not been defined unambiguously. An example is the various additional class I antigens mapped into the *H-2D* region (see Chapter One). Knowing how many genes are available in each region makes it now possible to transfect mouse cell lines with each gene and test to see whether these antigens can be detected now in the transfectants. Their possible origins from the defined class I genes via alternative RNA processing or posttranslational modifications can also be tested. The use of simple and highly specific test reagents, such as monoclonal antibodies and nucleic acid probes, will help to clarify the confusion left by classical serology.

Evolution and Polymorphism

As shown in Chapter Three, the class I gene family is a patchwork of multiple mutational events. These include point mutations, insertions, deletions, inversions, and gene conversions. The duplication and deletion of genes or blocks of genes probably are the result of homologous but unequal crossing over, and can serve to expand or contract the size of the gene family (see also Chapters One and Two).

Identification of these evolutionary processes provide clues to the mechanisms for generating the high degree of polymorphism of the H-2 class I antigens. In addition to the accumulation of point mutations, at least four recombinational mechanisms probably contribute to the generation and maintenance of the polymorphism. First, gene conversion appears to have a

primary role. This has been demonstrated most clearly in the studies with the H-2K^{bm} mutants (Pease et al., 1983; Mellor et al., 1983; Weiss et al., 1983) and in the present study (Chapter Three). Second, homologous recombination occurring within the genes can form a hybrid gene, thereby creating a new allele. The dm1 mutation is a clear example of this process (Chapter One). Third, homologous but unequal crossing-over events can expand or contract the size of the class I gene family. Since these genes can participate in gene conversions, they probably serve as a reserve for accumulating diversity. Fourth, homologous recombinations have been observed to occur with a high frequency at certain sites termed recombinational hot spots (Steinmetz et al., 1984; 1986). Recombination at the hot spots leads to the relatively rapid shuffling of chromosomal segments of the MHC and hence generation of new MHC haplotypes, and might be important for the maintenance of useful alleles during evolution.

REFERENCES

Amor, M., Tosi, M., Duponchel, C., Steinmetz, M., and Meo, T. (1985). Liver mRNA probes disclose two cytochrome P-450 genes duplicated in tandem with the complement *C4* loci of the mouse *H-2 S* region. Proc. Natl. Acad. Sci. USA. **82**, 4453-4457.

Chaplin, D. D., Woods, D. E., Whitehead, A. S., Goldberger, G., Colten, H. R., and Seidman, J. G. (1983). Molecular map of the murine *S* region. Proc. Natl. Acad. Sci USA. **80**, 6947-6951.

Fisher, D. A., Hunt S. W. III, and Hood, L. (1985). Structure of a gene encoding a murine thymus leukemia antigen, and organization of *Tla* genes in the BALB/c mouse. J. Exp. Med. **162**, 528-545.

Goodenow, R. S., McMillan, M., Nicolson, M., Sher, B. T., Eakle, K., Davidson, N., and Hood, L. (1982). Identification of the class I genes of the mouse major histocompatibility complex by DNA-mediated gene transfer. Nature **300**, 231-237.

Klein, J., Figueroa, F., and Nagy, Z. A. (1983). Genetics of the major histocompatibility complex: the final act. Ann. Rev. Immunol. **1**, 119-142.

Lalanne, J.-L., Transy, C., Guerin, S., Darche, S., Menlien, P., and Kourilsky, P. (1985). Expression of class I genes in the major histocompatibility complex: Identification of eight distinct mRNAs in DBA/2 mouse liver. Cell **41**, 469-478.

Mellor, A. L., Weiss, E. H., Ramachandran, K., and Flavell, R. A. (1983). A potential donor gene for the bm1 gene conversion event in the C57BL mouse. Nature **306**, 782-795.

Pease, L. R., Schulze, D. H., Pfaffenback, G. M., and Nathenson, S. G. (1983). Spontaneous H-2 mutants provide evidence that a copy mechanism analogous to gene conversion generates polymorphism in the major histocompatibility complex. Proc. Natl. Acad. Sci. USA **80**, 242-246.

Rogers, J. H. (1985). Family organization of mouse H-2 class I genes. Immunogenetics **21**, 343-353.

Steinmetz, M., Winoto, A., Minard, K., and Hood L. (1982). Clusters of genes encoding mouse transplantation antigens. Cell **28**, 489-498.

Steinmetz, M., Malissen, M., Hood, L., Orn, A., Maki, R.A., Dastoornikoo, G. R., Stephan, D., Gibb, E., and Romaniuk, R. (1984). Tracts of high or low sequence divergence in the mouse major histocompatibility complex. EMBO J. **3**, 2995-3003.

Steinmetz, M., Stephan, D., and Lindahl, K. F. (1986). Gene organization and recombinational hot spots in the murine major histocompatibility complex. Cell, in press.

Weiss, E., Golden, L., Zakut, R., Mellor, A., Fahrner, K., Kvist, S., and Flavell, R. A. (1983). The DNA sequence of the H-2K^b gene: Evidence for gene conversion as a mechanism for the generation of polymorphism in histocompatibility antigens. EMBO J. **2**, 453-462.

White, P. C., Chaplin, D. D., Weis, J. H., Dupont, B., New, M. I., and Seidman, J. G. (1984). Two steroid 21-hydroxylase genes are located in the murine *S* region. Nature **312**, 465-467.

Winoto, A., Steinmetz, M., and Hood, L. (1983). Genetic mapping in the major histocompatibility complex by restriction enzyme site polymorphisms: Most mouse class I genes map to the Tla complex. Proc. Natl. Acad. Sci. USA **80**, 3425-3429.