Analysis of Expression, Structure and Evolution of Non-classical Class I Major Histocompatibility Complex Genes

Thesis by

KURT A. BRORSON

In Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy

California Institute of Technology

Pasadena, California

1990

(Submitted September 7, 1989)

Acknowledgements

I would like first to thank Dr. Leroy Hood, my advisor and the other members of my committee, Drs. Ray Owen, Ellen Rothenberg, Giuseppe Attardi and David Anderson. I am also grateful to Mr. Tim Hunkapiller for his expert advice on nucleotide and protein sequence analysis; Drs. Hilde Cheroutre, Steve Hunt and Deborah Nickerson, for many years of collaboration; Drs. Joan Kobori and Iwona Stroynowski for insightful comments on several manuscripts, and Dr. Ben Koop for expert opinions on evolutionary analysis of DNA sequences. I am grateful to my parents, Carl and Connie Brorson, for their love and support through the years and for encouraging me to pursue a career in science. I am grateful for the friendship and encouragement of Jaime Gonzalez, Raymundo Madrid, Steve Tankersly and Chris Ware. I would like to thank Drs. Michel Klein and Robin Lowry, for their friendship and encouragement, and for being themselves. I want to thank Drs. Martha Zuniga and Kirsten Fischer Lindahl for taking an early interest in me, and for providing encouragement at critical points in my graduate tenure. I am grateful to Mrs. Cathy Blagg for expert secretarial assistance, and to Ms. Chin-Sook Kim, Elaine Gese, Karyl Minard, Barbara Otto, Teresa Geffroy, Annette Yuen, Bertha Jones, Jessie Walker, Maria DeBruyn, Tippy Harlow, Karen Gregorich and Doreen Harcus for making the lab a pleasant place to work.

Abstract

Class I major histocompatibility molecules (MHC) are 45 kilodalton (kD) glycoproteins that associate with a smaller 12 kD polypeptide, β_2-microglobulin. In the BALB/c mouse, there are three classical class I molecules, H-2K^d, D^d, and L^d, which are expressed throughout the body and present viral antigens to cytotoxic T lymphocytes (CTLs). In addition to the genes that encode the three classical class I antigens, the BALB/c genome contains 32 genes that structurally resemble the classical class I genes, and therefore possibly encode class I molecules. A few of the non-classical class I genes have been shown to encode molecules, TL, Qa-1, Qa-2, Q10, Qb-1, and Hmt, which are expressed in a generally tissue-specific manner, and probably do not act as restriction elements. However, it is unclear what function these molecules play, or why such a large gene family is maintained if only three viral antigen-presenting restriction elements are required by the murine immune system.

DNA sequences were obtained from each of the 35 class I genes of the BALB/c mouse of the transmembrane domain-encoding fifth exon. Based on nucleotide sequence similarity, the fifth exons could be divided into seven groups that share little similarity with each other. In addition, the majority of the fifth exons are able to encode a transmembrane domain that can be separated into a proline-rich connecting peptide, a hydrophobic transmembrane segment, and a ctyoplasmic portion that includes basic anchoring residues. Since this conservation occurs in spite of extensive variation of nucleotide sequence in these exons, it is likely that selective pressure exists to maintain a functional structure in the majority of class I genes.

A cDNA library was constructed from a thymus from a five-week-old BALB/c mouse. From this library, 69 class I cDNA transcripts from 15 different

class I genes were isolated and analyzed. Included were three novel transcripts from *Tla* subregion genes, the $T9^c$, $T17^c$, and $T18^c$ genes. Sequence analysis of these clones reveals that the $T9^c$ gene is probably a pseudogene, while the $T18^c$ gene has an open reading frame in at least exons 2, 3, 4, and 5. A fourth cDNA clone was a transcript from the *Thy19.4* gene, a gene that had not been previously isolated on a recombinant DNA clone. The isolation of transcripts from such a relatively large number of genes suggests that the number of expressed and perhaps functionally important class I genes may be larger than previously believed, and that expression of class I recognition structures may be important for cell-cell interactions within the thymus.

To further pursue the characterization of the *Thy19.4* gene, a genomic clone containing this gene was isolated from a size-selected insert library, and the DNA sequence of the *Thy19.4* gene was obtained. The *Thy19.4* gene contains an open reading frame, and in several aspects resembles the genes that encode the transplantation antigens. These similarities include a shared exon/intron structure and shared amino acid sequence motifs. In addition, PCR amplification experiments using tissue cDNA demonstrates that the *Thy19.4* gene is expressed in a variety of tissues. However, unlike the classical transplantation antigens, the *Thy19.4* gene maps distal to the *H-2* region, in the *Hmt* region.

These studies have demonstrated that class I gene transcription is more extensive than previously believed. Some of the expressed genes, like the $T18^c$ and *Thy19.4* genes, appear to be able to encode class I molecules which may share structural characteristics with the classical transplantation antigens and may possibly serve as recognition structures in cell-cell interaction events. In addition, examination of the transmembrane domain exon of each of the 35 class I genes suggests that some selective constraint is acting on the majority of

members of this family of genes, thus raising the possibility that many of the non-classical class I genes encode functionally important products.

Table of Contents

INTRODUCTION

The Immune System. Mammalian immunity is a system of cellular and humoral components that identify and eliminate foreign pathogens (Klein 1982). The humoral component of the immune system includes a variety of serum proteins, which perform several functions. Immunoglobulins are glycoproteins that specifically bind to and neutralize foreign pathogens. Complement is a group of proteins that together mediate a variety of functions, including lysis of cellular targets. The interleukins are synthesized by the cells of the immune system and modulate their behavior in a variety of ways. The cells of the immune system include monocyte/macrophages, granulocytes, and lymphocytes. Monocytes, macrophages, and granulocytes mediate non-specific effector functions such as antigen presentation or phagocytosis, while antigen-specific functions are mediated by lymphocytes. There are two types of lymphocytes denoted T cells and B cells. Upon activation, B cells produce antibodies. T-helper lymphocytes mediate the differentiation of T- and B-cell responses by secreting interleukins. Cytotoxic T-lymphocytes (CTLs) recognize and destroy virus-infected and tumor cells.

Lymphocytes have antigen-specific receptors on their cell surfaces through which they are activated by antigens. These receptors are immunoglobulins on B cells and T-cell receptors (TCRs) on T cells. An antigen alone can stimulate B cells, although they require concurrent T-cell help in the form of interleukins. T cells, on the other hand, are stimulated by a peptide fragment of the antigen only when it is presented by an accessory cell. This phenomenon occurs because the ligand of the T-cell receptor is the peptide epitope antigen in physical association with a second receptor component, a molecule encoded by the major histocompatibility complex (MHC) (Zinkernagel and Doherty 1980).

Classical Major Histocompatibility Complex Molecules. There are two types

of MHC molecules: class I and class II (Klein 1986; Figure 1). Class I MHC molecules are 45 kilodalton (kd) cell-surface glycoproteins that associate non-covalently with a small polypeptide, β_2-microglobulin. There are two types of class I molecules, classical and non-classical, which differ in function and expression. In mice the three classical class I molecules are named K, D, and L (Table I), while in humans they are named A, B, and C. Classical class I molecules are expressed on almost all somatic cells with the exception of the brain. Class II molecules, on the other hand, are heterodimer glycoproteins consisting of a 33 kd α chain, and a 29 kd β chain. The two class II molecules of mice are named I-A, and I-E, while the three class II molecules of humans are named DP, DQ, and DR. The class II molecules have a more restricted range of tissue distribution than class I molecules. They are expressed on macrophages, B cells, and other antigen-presenting cells. The classical MHC molecules display a high degree of polymorphism within a species, and were thus first detected either serologically or by transplant rejection reactions (Gorer 1936). Classical class I molecules present viral antigens to CD8-expressing T cells, usually CTLs. On the other hand, class II molecules present peptide antigens to T-helper cells, usually after the antigen has been processed intracellularly by the antigen-presenting cell, and thus are involved in the initiation of antigen-specific immune responses.

The structure of MHC molecules, which has been determined for the human class I molecule HLA-A2, is critical for their function as antigen-presenting structures (Bjorkman *et al.* 1987a and b; Figure 2). Class I glycoproteins have five protein domains, three external domains $\alpha 1$, $\alpha 2$, and $\alpha 3$, as well as a transmembrane and a cytoplasmic domain (Kabat *et al.* 1977). Each of the external domains is approximately 90 amino acids in length. The structure of the $\alpha 3$ domain as well as of β_2-microglobin is an immunoglobulin homology-unit

domain. These homology units interact closely in the class I molecule. The α1 and α2 domains together form a β-pleated sheet and two α helices that rest above the β-pleated sheet platform. Together these two α helices form a cleft in which viral antigen peptides are proposed to reside during antigen presentation. The hydrophobic transmembrane domain is about 24 amino acids in length, while the cytoplasmic domain is usually around 40 amino acids long.

Non-classical Class I Major Histocompatibility Complex Molecules. In addition to the classical antigen-presenting class I molecules, several other mouse class I molecules have been detected either serologically or with cloned CTL lines (Table I). These non-classical class I molecules resemble classical class I molecules structurally, but display more limited polymorphism, have restricted tissue distribution, and do not have known functions. The first of these non-classical class I molecules to be characterized is the TL, or thymus-leukemia antigen. The TL antigen was originally detected by an antiserum produced in B6 mice against A-strain leukemia cells (Old ***et al.*** 1963). It was later found to be expressed on small cortical thymocytes (Fischer Lindahl and Langhorne 1981; Rothenberg 1982), as well as on peripheral T cells activated by concanavalin A (Cook and Landolfi 1983). Exactly five variants of the TL antigen have been identified by a panel of monoclonal antibodies, while some mice, such as the B6 mouse used to produce TL antiserum, do not express TL antigens (Shen ***et al.*** 1982).

A second non-classical class I molecule, the Qa-1 antigen, was identified using the antisera originally used to define the TL antigen (Stanton and Boyse 1976). The Qa-1 antigen is dissimilar from the TL antigen in that it is expressed in a variety of tissues. In addition, it can be detected with anti-TL antiserum that has been absorbed against thymocytes, and hence is structurally distinct from the

TL antigen. Anti-Qa-1 CTL lines have been generated (Aldrich *et al.* 1988). Thus the antigen itself is capable of acting as a restriction element, at least for itself.

The Qa-2 antigen was discovered when recombinant inbred strains of mice were produced with recombination breakpoints between the genes that encode the H-2 and TL antigens (Flaherty 1976). An antiserum was produced by immunizing one of these recombinant mice, B6.K1, with B6 spleen and lymph-node cells. This antiserum detected a class I element whose gene mapped between the breakpoints in two of these mice, B6.K1, and B6.K2. This class I molecule, the Qa-2 antigen, is expressed in a variety of tissues, including thymus, spleen, and lymph node. Interestingly, this molecule is not fixed to the cell surface by a classical transmembrane domain, but instead is linked by a phosphatidylinositol linkage (Stroynowski *et al.* 1987). There are two variants of the Qa-2 molecule, the high- and low-expressed variants.

Two secreted class I molecules have been described, the Qb-1 molecule (Robinson 1985), and the Q10 molecule (Maloy *et al.* 1984). These two class I molecules are secreted because they lack either a hydrophobic transmembrane domain, or anchoring residues at the end of the transmembrane domain. Like many other serum proteins, the Q10 molecule is synthesized in the liver, and can reach serum concentrations of 60 μg/ml. On the other hand, the Qb-1 antigen gene is transcribed in a variety of tissues. There are two variants of Qb-1, an acidic (a), and a basic (b) form, while some strains of mice do not express the Qb-1 molecule. Since both of these molecules lack a cytoplasmic domain, they are smaller than H-2 molecules; Qb-1 is 41 kd, while Q10 is 40 kd.

Hmt is a non-classical class I molecule that was discovered because it is a component of an unusual CTL target antigen, the maternally transmitted antigen (Mta). Mta is a complex of the Hmt class I molecule, and a short mitochondrially

derived peptide (Mtf). The Mta antigen was originally detected by CTL lines generated by immunizing NZB/B1NJ mice with *H-2* compatible BALB/c cells (Fischer Lindahl *et al.* 1980). These CTLs seemingly defied MHC restriction since they killed target cells from 80 different strains of mice, regardless of the *H-2* type. Most surprisingly, crosses of Mta^+ mice with NZB mice demonstrated that the antigenic determinant was inherited maternally, suggesting a mitochondrial origin. Four variants of the mitochondrial factor have been found: α, β, γ, and δ. Subsequently, the mitochondrial determinant was identified as the leader segment of the ND1 protein (K. Fischer Lindahl, personal communication). The discovery that Mta is not expressed in ***Mus musculus castaneus*** mice, which have the α variant mitochondria, revealed that a chromosomal locus, designated *Hmt*, that maps near to, but not within, the *H-2* locus is also required to form the Mta antigen (Fischer Lindahl *et al.* 1987). Studies with β_2-microglobulin$^-$ cell lines revealed that the protein encoded within the chromosomal locus require β-microglobulin coexpression, and thus is probably a non-classical class I molecule. There are three variants of Hmt: a, c, and d. The a variant is that of ***domesticus*** mice, while ***Mus musculus castaneus*** mice do not express the Hmt element. The Hmt element is similar in many ways to the classical class I antigens. It is expressed in almost all somatic tissues, and appears to be able to present at least one antigen to T cells, the Mtf factor. It is probably fortuitous that an endogenous mitochondrial peptide is presented by this class I molecule. However, this observation does suggest that the Hmt molecule may function as a presenting structure for antigens localized in subcellular components that transplantation antigens do not normally traffic through.

Possible Functions of Non-classical Class I Molecules. The functions of the non-classical class I antigens are not known, but the wide variety of them suggests

that they probably have a multiplicity of functions. Initially, these molecules were proposed to be cell type-specific recognition structures since some are relatively non-polymorphic and seem to have a restricted cell-surface tissue distribution (Fischer Lindahl and Langhorne 1981; Williams 1982). The conservation of critical amino acid residues of the presumed products of some class I genes such as *Thy19.4* discussed in Chapter 3 suggests that some may resemble structurally the transplantation antigens and may possibly be recognition structures involved in cell-cell interaction events. Chapter 2 demonstrates that many of the non-classical class I genes are evolutionarily ancient, and hence, may encode recognition structures whose function is common to a variety of vertebrates, such as that required for development. The best evidence supporting the contention that class I molecules are involved in development is the finding that a variation in blastula cleavage rate in mice is linked to the *Qa-2* gene, suggesting that this gene or another tightly linked gene is involved in early development (Warner *et al.* 1987).

A second possible function of some of these molecules is that they are antigen-presenting structures for a new class of T-cell receptors, the $\gamma\delta$ receptors (Strominger 1989). The $\gamma\delta$ T-cell receptors have a limited range of diversity because of the small number of variable region gene segments that they employ. Hence, the relatively low polymorphism of the non-classical class I antigens would be evolutionarily advantageous if they were ligands for $\gamma\delta$ T-cell receptors. So far, very few $\gamma\delta^+$ T-cell hybridomas have been isolated that are specific for non-classical class I molecules, and further experiments will be needed to determine whether these hybridomas are typical of $\gamma\delta^+$ T cells (Bluestone *et al.* 1988; S. Tonegawa, personal communication).

Another possible function involving immune recognition is raised by the Mta

antigen (K. Fischer Lindahl, personal communication). Since this antigen presents a mitochondrial peptide, it is possible that it and perhaps other non-classical class I molecules present antigens from subcellular components that the classical H-2 molecules do not traffic through. Clearly, when the *Hmt* gene and its protein product are isolated and characterized, experiments can be conducted to trace the trafficking of this class I molecule and to test this possibility.

The soluble class I molecules, Q10 and Qb-1, probably serve entirely different functions than the other non-classical class I molecules. Suggestions that they may be involved with tolerance to self-MHC have proven to be unlikely (Mann *et al.* 1987). However, the intriguing finding that responses to odors by mice are encoded by genes that map to the *Qa/Tla* region suggests that some non-classical class I molecules may have an entirely non-immunologically related function (Yamazaki *et al.* 1982). Since these soluble class I molecules are present at relatively high concentrations in the serum, it is conceivable that they may be secreted into the urine and may function as chemosensory identification signals (Maloy *et al.* 1984).

The MHC Gene Loci. In the mouse, there are between 28 and 35 class I genes (Weiss *et al.* 1984; Steinmetz *et al.* 1982; Figure 3). These class I genes map to five subregions defined by recombinant mice: *K*, *D*, *Qa*, *Tla*, and *Hmt* (Winoto *et al.* 1983; Richards *et al.* 1989). The *H-2* region is composed of the *K* and *D* subregions along with the *I* subregion, which contains class II genes, and the *S* subregion, which contains complement component genes. In the BALB/c mouse there are two genes in the *K* region, K^d, and $K2^d$, while there are five genes in the *D* region: D^d, $D2^d$, $D3^d$, $D4^d$, and L^d. The three classical class I antigens are encoded by the K^d, D^d, and L^d genes.

The BALB/c *Qa* subregion contains eight class I genes: $Q1^d$, $Q2^d$, $Q4^d$, $Q5^d$,

$Q6^d$, $Q7^d$, $Q8/9^d$, and $Q10^d$. The nomenclature of these genes corresponds to that of the C57BL mouse, which has ten *Qa* region genes: $Q1^b$ through $Q10^b$ (Weiss *et al.* 1984). The gene corresponding to the $Q3^b$ gene was lost in the evolution of the BALB/c MHC, while the genes that corresponded to the $Q8^b$ and $Q9^b$ genes fused during an unequal crossover event, and thus became the hybrid gene $Q8/9^d$ of the BALB/c mouse. A large number of non-classical class I molecules are encoded by genes within this subregion. The Qa-2 molecule is encoded by the $Q7^d$ gene (Stroynowski *et al.* 1987), the Qb-1 molecule is encoded by the $Q4^d$ gene (Robinson *et al.* 1988), while the Q10 molecule is the product of the $Q10^d$ gene (Cosman *et al.* 1982).

The *Tla* subregion of the BALB/c mouse has at least 19 class I genes: $T1^c$ through $T18^c$, and the 37^c gene, which resides adjacent to the $T17^c$ gene (Fisher *et al.* 1985; Transy *et al.* 1987). These genes map to one of two clusters containing $T1^c$ through $T10^c$ and $T11^c$ through 37^c. These two clusters were created in a gene-block duplication event. The $T1^c$, $T2^c$, and $T3^c$ genes are duplicates of the $T11^c$, $T12^c$, and $T13^c$ genes, while the $T6^c$ through $T10^c$ genes are duplicates of the $T14^c$ through 37^c genes. The $T4^c$ and $T5^c$ genes are proposed to have been created in a duplication event that also created the $T6^c$ and $T7^c$ genes. However, data presented in this thesis (Chapter 1) suggest that it is more likely that the $T4^c$ and $T5^c$ genes are partially duplicated class I genes derived from separate sources. The $T18^c$ gene does not map to either of these two clusters, and exists as a single copy in the BALB/c genome. The TL antigen is encoded within this subregion by the $T13^c$ gene (Fisher *et al.* 1985), while the 37^c gene, which has an open reading frame, is transcriptionally active but has not been shown to encode a protein (Transy *et al.* 1987).

The *Hmt* region is a newly described subregion (Richards *et al.* 1989). At

this point, it is clear that there are at least three class I genes in this subregion, including the gene that encodes the Hmt element of the maternally transmitted antigen. In addition, the *Thy19.4* gene, a new class I gene described in this thesis (Chapter 3), also maps to this region. Finally, an additional mouse class I gene, *Mb-1*, has been isolated that has not yet been mapped to any of these subregions (Singer *et al.* 1988).

Structure of Class I Genes. Class I genes contain between 5 and 8 exons (Steinmetz *et al.* 1981; Singer *et al.* 1988; Figure 4). Exon 1 encodes an approximately 30 amino acid hydrophobic leader segment which presumably assists in the transport of the glycoprotein, and is cleaved posttranslationally. Exons 2, 3, and 4 encode the approximately 90 amino acid $\alpha 1$, $\alpha 2$, and $\alpha 3$ external domains. Exon 4 is the most conserved exon of class I genes. It is probably conserved to maintain the tertiary structure of the $\alpha 3$ domain that is required for interaction with the virtually invariant β_2-microglobulin molecule. Exon 5 encodes the hydrophobic transmembrane domain. The translation of the fifth exon includes a short connecting peptide, an approximately 23 amino acid transmembrane segment, and 3-4 basic residues that presumably assist in anchoring the class I molecule to the membrane. Finally, the short cytoplasmic domain is encoded by exons 6, 7, and 8.

Significance of the Size of the Class I Gene Family. There are at least 35 class I genes in the BALB/c mouse, and only nine characterized class I molecules. Only three of these genes, K^d, D^d, and L^d, encode transplantation antigens, while the other six molecules have no known function. It is unclear why the mouse genome contains 35 class I genes when three transplantation antigens are sufficient to present viral antigens to CTLs. However, it is unlikely that such a large family of genes would be maintained in several species (Steinmetz *et al.*

1982; Srivastava *et al.* 1985) if the non-classical class I genes had no function. Other families of genes such as the β-globin genes (Edgell *et al.* 1985) and the immunoglobulin V genes (C. Readhead, personal communication) have pseudogenes, but generally there is not a preponderance of them. At this point, very few class I genes have been characterized on any level, particularly those in the *Tla* region (Table II). Aside from those shown to encode proteins, a few pseudogenes have been discovered among them. The $T1^c$ and *T2* genes are pseudogenes since they have several frameshifts and stop codons in their exons, and hence are incapable of encoding class I proteins (Widmark *et al.* 1988; Fisher *et al.* 1989). The $T5^c$ (Rogers 1985), $T10^c$ (D. Nickerson, personal communication), and $T13^b$ genes (Brown *et al.* 1988) all have a stop codon in at least one of their exons and are thus probably pseudogenes. However, it is possible that in some cases transcripts from these genes could use alternative RNA splicing patterns to avoid placing the stop codons in the mRNA reading frame, and thus remain functional. This thesis adds $T9^c$ to the list of pseudogenes (Chapter 2).

Thesis. To initiate a study of the family of non-classical class I genes, I have employed a variety of approaches to determine which genes are pseudogenes, which are expressed (Chapter 2), and if they are expressed, whether they can encode class I molecules (Chapters 2, 3 and Appendix I). Sequence analysis of entire genes (Chapter 3) or portions of several genes (Chapters 1 and 2) can predict structural aspects of the putative molecules encoded by these genes, and thus possibly hint at their function. Finally, protein expression analysis was initiated by the construction of transfectants for several *Tla* region genes (Appendix I).

The first chapter of this thesis describes the sequencing and analysis of the fifth exons of the 35 BALB/c class I genes. This study was initiated to allow the

identification of class I cDNA transcripts by comparison of exon 5 sequences. In the analysis of the sequences, it was found that the majority of these fifth exons are able to encode a transmembrane domain that can be separated into a proline-rich connecting peptide, a hydrophobic transmembrane segment, and a cytoplasmic portion that includes basic anchoring residues. This conservation occurs despite extensive variation of nucleotide sequence in these exons, suggesting that selective pressure exists in the majority of class I genes to maintain a functional structure.

The second chapter describes the isolation and characterization of class I transcripts from a cDNA library constructed from five-week-old BALB/c thymus. Sixty-nine class I cDNA clones were isolated from this library, and among them were transcripts from 15 different class I genes. Included were three previously undescribed transcripts from *Tla* subregion genes, $T9^c$, $T17^c$, and $T18^c$. Sequence analysis of these clones reveals that the $T9^c$ gene is a pseudogene. A $T17^c$ clone was also partially sequenced, and exhibited open reading frames as far as examined. The $T18^c$ clone had an open reading frame in at least exons 2, 3, 4, and 5 and hence could probably encode a class I protein. A fourth cDNA clone was isolated from the *Thy19.4* gene, a gene that had not been previously isolated molecularly. This gene became the focus of the studies described in Chapter 4. In addition, a clone from the $D2^d$ gene was characterized, a gene that became the object of study in another lab (Hedley *et al.* 1989). The isolation of transcripts from a relatively large number of genes suggests that the number of expressed and functionally important class I genes may be larger than previously believed. Since the majority of the transcripts examined have open reading frames, it is likely that the class I genes expressed in the thymus are important for the function of that tissue.

Chapter 3 describes a novel class I gene, *Thy19.4*, which was initially isolated from the thymus cDNA library. A genomic clone containing this gene was isolated from a size-selected insert library and sequenced. The *Thy19.4* contains an open reading frame, and in several aspects resembles the genes that encode transplantation antigens. These similarities include a shared exon/intron structure and amino acid sequence motifs. In addition, PCR amplification experiments using tissue cDNA demonstrates that the *Thy19.4* gene is expressed in a variety of tissues. However, unlike the classical transplantation antigens, the *Thy19.4* gene maps distal to the *H-2* region, in the *Hmt* region.

Finally, Appendix I describes an attempt to construct a series of transfectants for the *Tla* region class I genes. Each transfectant has only one *Tla* region gene transfected into it, and the parental line into which the *Tla* genes were transfected is from a B10.CAS2 mouse, which has the $H\text{-}2^{w17}$, Tla^{g}, and Hmt^{b} haplotypes of the ***Mus musculus castaneus*** subspecies of mouse. Because of this, the probability of allelism between the transfected Tla^{c} region gene and the endogenous Tla^{g} genes of the parental cell line is higher than if a ***Mus musculus domesticus*** cell line is used. These transfectants will, it is hoped, prove to be useful for protein characterization of *Tla* region gene products.

In this thesis, I have attempted to initiate studies on the non-classical class I genes, with a particular emphasis on the *Tla* region genes on which very few studies have been performed. It is my hope that the groundwork studies presented in this thesis will provide a clear basis for further studies of the *Tla* region genes, and the non-classical class I genes in general.

References

Aldrich, C., Rodgers, J., and Rich, R. Regulation of Qa-1 expression and determinant modification by an *H-2D*-linked gene, ***Qdm***. *Immunogenetics* **28**:334-344, 1988.

Bjorkman, P., Saper, M., Samraoui, B., Bennett, W., Strominger, J., and Wiley, D. Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* **329**:506-512, 1987a.

Bjorkman, P., Saper, M., Samraoui, B., Bennett, W., Strominger, J., and Wiley, D. The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature* **329**:512-518, 1987b.

Bluestone, J., Cron, R., Cotterman, M., Houlden, B., and Matis, L. Structure and specificity of T cell receptor $\gamma\delta$ on major histocompatibility complex antigen-specific $CD3^+$, $CD4^-$, $CD8^-$ T lymphocytes. *J. Exp. Med.* **168**:1899-1916, 1988.

Brown, G., Choi, Y., Egan, G., and Meruelo, D. Extension of the *H-2* Tla^b molecular map. *Immunogenetics* **27**:239-251, 1988.

Cook, R., and Landolfi, W. Expression of the thymus leukemia antigen by activated peripheral T lymphocytes. *J. Exp. Med.* **158**:1012-1017, 1983.

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

Edgell, M., Hardies, S., Brown, B., Voliva, C., Hill, A., Phillips, S., Cormer, M., Burton, F., Weaver, S., and Hutchinson, C. Evolution of the mouse β globin complex locus. In: *Evolution of Genes and Proteins*, Nei, M., and Koehn, R., eds. Shinaver Associates, Inc., Sunderland, MA, 1985.

Fischer Lindahl, K., Boccheieri, M., and Riblet, R. Maternally transmitted target antigen for unrestricted killing by NZB T lymphocytes. *J. Exp. Med.* **152**:1583-1595, 1980.

Fischer Lindahl, K., and Langhorne, J. Medial histocompatibility antigens. *Scand. J. Immunol.* **14**:643-654, 1981.

Fischer Lindahl, K., Loveland, B., and Richards, C. The end of *H-2*. In: *Major Histocompatibility Genes and Their Roles in Immune Function.* David, C. S., ed. Plenum Press, NY, 1987.

Fisher, D., Hunt, S., and Hood, L. 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, 1985.

Fisher, D., Pecht, M., and Hood, L. DNA sequence of a class I pseudogene from the *Tla* region of the murine MHC: Recombination at a B2 Alu repetitive sequence. *J. Mol. Evol.* **28**:306-312, 1989.

Flaherty, L. The *Tla* region of the mouse: Identification of a new serologically defined locus, *Qa-2*. *Immunogenetics* **3**:533-539, 1976.

Gorer, P. The detection of a hereditary antigenic difference in the blood of mice by means of a human group A serum. *J. Genetics* **32**:2-31, 1936.

Hedley, M., Hunt, S., Brorson, K., Andris, J., Tucker, P., Hood, L., and Forman, J. DNA sequence analysis of $D2^d$: A new *D* region class I gene. *Immunogenetics* **29**:359-365, 1989.

Hunkapiller, T., and Hood, L. Diversity of the immunoglobulin gene superfamily. *Adv. Immunology* **44**:1-63, 1989.

Kabat, E., Wu, T., Reid-Miller, M., Perry, H., and Gottesman, K. *Sequences of proteins of immunological interest*, 4th ed. U.S. Department of Health and Human Services, Public Health Services, National Institutes of Health, 1977.

Klein, J. *Immunology: The Science of Self-Nonself Discrimination.* John Wiley and Sons, New York, 1982.

Klein, J. *Natural History of the Major Histocompatibility Complex.* John Wiley and Sons, New York, 1986.

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

Maloy, W., Coligan, J., Barra, Y., and Jay, G. 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, 1984.

Mann, D., Stroynowski, I., Hood, L., and Forman, J. Cytotoxic T lymphocytes from mice with soluble class I Q10 molecules in their serum are not tolerant to membrane bound Q10. *J. Immunol.* **138**:240-245, 1987.

Moore, K., Sher, B., Sun, Y., Eakle, K., and Hood, L. DNA sequences of a gene encoding a BALB/c mouse L^d transplantation antigen. *Science* **215**:679-682, 1982.

Old, L., Boyse, E., and Stockert, E. Antigenic properties of experimental leukemias. I. Serological studies in vitro with spontaneous and radiation induced leukemias. *J. Natl. Cancer Inst.* **31**:977-986, 1963.

Richards, S., Bucan, M., Brorson, K., Kiefer, M., Hunt, S., Lehrach, H., and Fischer Lindahl , K. Genetic and molecular mapping of the *Hmt* region of the mouse. *EMBO J.*, 1989, in press.

Robinson, P. Qb-1, a new class I polypeptide encoded by the *Qa* region of the mouse *H-2* complex. *Immunogenetics* **22**:285-289, 1985.

Robinson, P., Bever, D., Mellor, A., and Weiss, E. Sequence of the mouse *Q4* class I gene and characterization of the gene product. *Immunogenetics* **27**:79-86, 1988.

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

Rothenberg, E. A specific biosynthetic marker for immature thymic lymphoblasts: Active synthesis of thymus leukemia antigen restricted to proliferating cells. *J. Exp. Med.* **155**:140-154, 1982.

Shen, F., Chorney, M., and Boyse, E. Further polymorphism of the *Tla* locus defined by monoclonal TL antibodies. *Immunogenetics* **15**:573-578, 1982.

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

Singer, D., Hare, J., Golding, H., Flaherty, L., and Rudikoff, S. Characterization of a new subfamily of class I genes in the *H-2* complex of the mouse. *Immunogenetics* **28**:13-21, 1988.

Srivastava, R., Duceman, B., Biro, D., Sood, A., and Weissman, S. Molecular organization of the class I genes of the human major histocompatibility complex. *Immunol. Rev.* **85**:93-121, 1985.

Stanton, T., and Boyse, E. A new serologically defined locus, *Qa-1*, in the *Tla* region of the mouse. *Immunogenetics* **3**:525-531, 1976.

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

Steinmetz, M., Winoto, A., Minard, K., and Hood, L. Clusters of genes encoding mouse transplantation antigens. *Cell* **28**:489-492, 1982.

Strominger, J. The $\gamma\delta$ T cell receptor and class Ib MHC-related proteins: enigmatic molecules of immune recognition. *Cell* **57**:895-898, 1989.

Stroynowski, I., Soloski, M., Low, M., and Hood, L. A single gene encodes soluble and membrane bound forms of the major histocompatibility Qa-2 antigen: Anchoring of the product by a phospholipid tail. *Cell* **50**:759-768, 1987.

Transy, C., Nash, S., David-Watine, B., Cochet, M., Hunt, S., and Hood, L. A low polymorphic mouse *H-2* class I gene from the *Tla* complex is expressed in a broad variety of cell types. *J. Exp. Med.* **166**:341-361, 1987.

Warner, C., Gollnick, S., Flaherty, L., and Goldbard, S. Analysis of Qa-2 expression by preimplantation mouse embryos: Possible relationship to the preimplantation embryo-development (*Ped*) gene product. *Biol. Reprod.* **36**:611-616, 1987.

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

Widmark, E., Ronne, H., Hammerling, U., Servenius, B., Larhammar, D., Gustafsson, K., Bohme, J., Peterson, P., and Rask, L. Family relationships of murine major histocompatibility complex class I genes. Sequence of the $T2A^{a}$ pseudogene, a member of gene family 3. *J. Biol. Chem.* **263**:7055-7059, 1988.

Williams, A. Surface molecules and cell interactions. *J. Theor. Biol.* **98**:221-234, 1982.

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

Yamazaki, K., Beauchamp, G., Bard, J., Thomas, L., and Boyse, E. Chemosensory recognition of phenotypes determined by the *Tla* and *H-2K* regions of chromosome 17 of the mouse. *Proc. Natl. Acad. Sci. USA* **79**:7828-7831, 1982.

Zinkernagel, R., and Doherty, P. MHC-restricted cytotoxic T cells: studies on the role of polymorphic major transplantation antigens determining T-cell restriction specificity, function, and responsiveness. *Adv. Immunol.* **27**:51-177, 1980.

Table I

Class I Molecules

Molecule	Expression	Functions	Reference
K^d, D^d, L^d	Entire body	Restriction elements	Klein 1986
Qa-1	Entire body	?	Stanton and Boyse 1976
Qa-2	Lymphoid cells	?	Flaherty 1976
TL	Thymus	?	Old *et al.* 1963
Q10	Liver-soluble	?	Maloy *et al.* 1984
Qb-1	Entire body - soluble	?	Robinson 1985
Hmt	Entire body	?	Fischer Lindahl *et al.* 1980

Table II

Status of Class I Genes

Gene	Product	Expression	Reference
K^d	H-2K^d	Entire body	Kvist *et al.* 1983
D^d	H-2D^d	Entire body	Sher *et al.* 1985
L^d	H-2L^d	Entire body	Moore *et al.* 1982
$D2^d$	?	?	Hedley *et al.* 1989
Q4	Qb-1 soluble class I	Entire Body	Robinson *et al.* 1988
$Q7^d$	Qa-2 antigen	Lymphoid	Steinmetz *et al.* 1981
Q10	Q10 soluble class I	Liver	Cosman *et al.* 1982
$T13^c$	TL antigen	Thymus	Fisher *et al.* 1985
37^c	?	Entire body	Transy *et al.* 1987
$T1^c$	ψ gene	-	Fisher *et al.* 1989
T2	ψ gene	-	Widmark *et al.* 1988
$T5^c$	ψ gene	-	Rogers 1985
$T10^c$	ψ gene	-	D. Nickerson, personal communication

The ***Q4***, ***Q10*** and ***T2*** genes were characterized in other strains of mice but are presumably similar in BALB/c mice.

Figure 1. ***Schematic diagrams of class I and II MHC molecules.*** Immunoglobulin homology unit domains are indicated as circular loops, while other domains are shown as wavy loops. Carbohydrates present on some MHC molecules are indicated as jagged lines. Disulfide bonds are shown as pairs of the letter S. The lipid bilayer is shown as two horizontal lines enclosing schematic phospholipids. Figure adapted from Hunkapiller and Hood (1989).

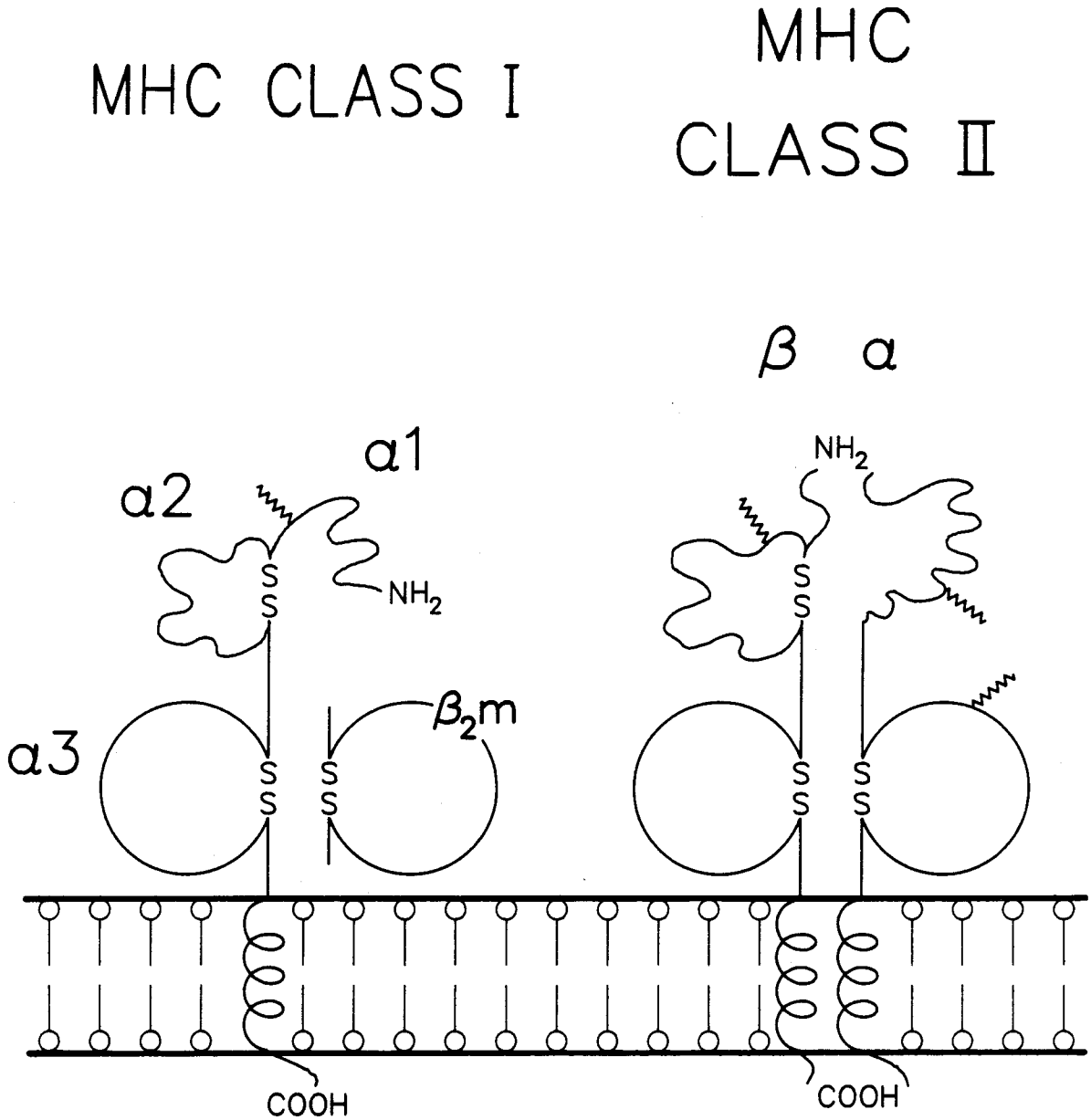

MHC CLASS I
MHC CLASS II
β α
α1
α2
α3
NH2
β2m
S S
COOH

Figure 2. *Schematic representation of the tertiary structure of the external domains of the human HLA-A2 class I MHC molecule.* β-strands are shown as thick arrows pointing in the amino-to-carboxy direction, and α helices are represented as helical ribbons. Connecting loops are depicted as thin lines, while disulfide linkages are shown as two spheres connected by lines. Figure adapted from Bjorkman *et al.* (1987a).

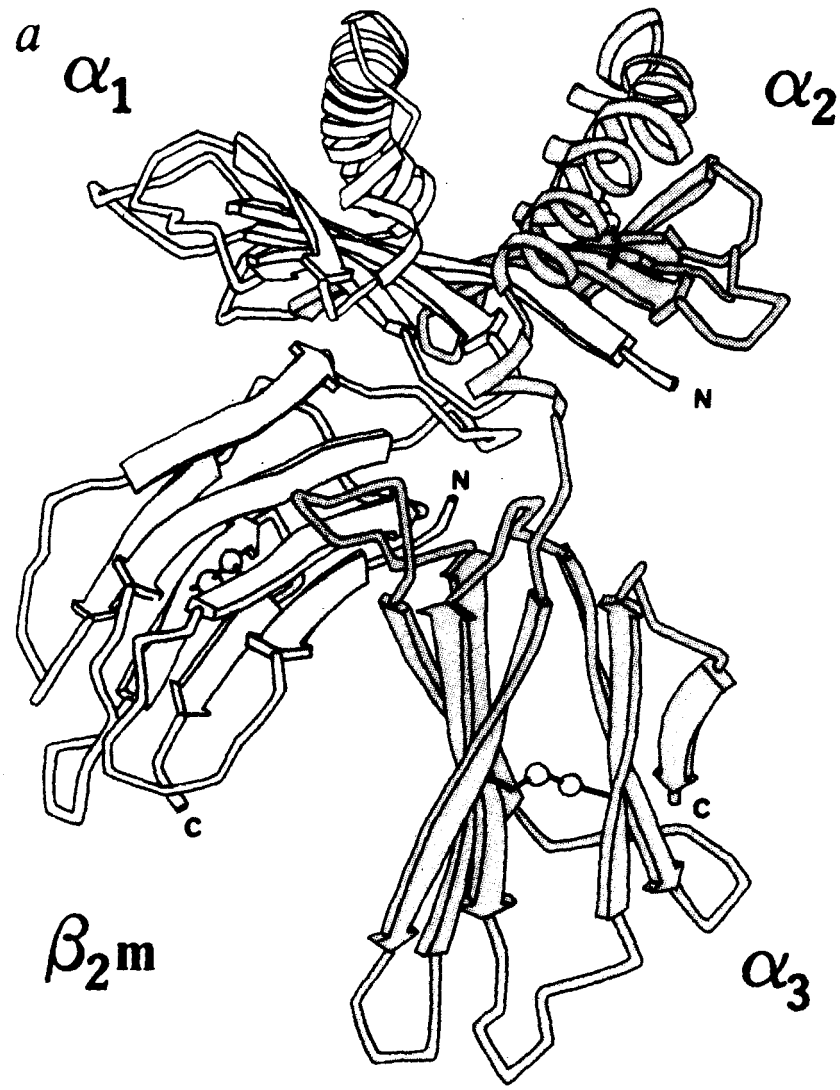

a
α1
α2
N
N
C
C
β2m
α3

Figure 3. ***Map of class I genes in the BALB/c MHC.*** Class I genes map to the *K*, *D*, *Qa*, *Tla*, and *Hmt* regions. The *I* and *S* regions contain class II and complement genes, respectively. The order of the *Tla* region gene clusters is unknown, as is the distance between the *K*, *D*, *Tla*, and *Hmt* regions. The upper line represents the genetic map and the gene clusters are indicated below.

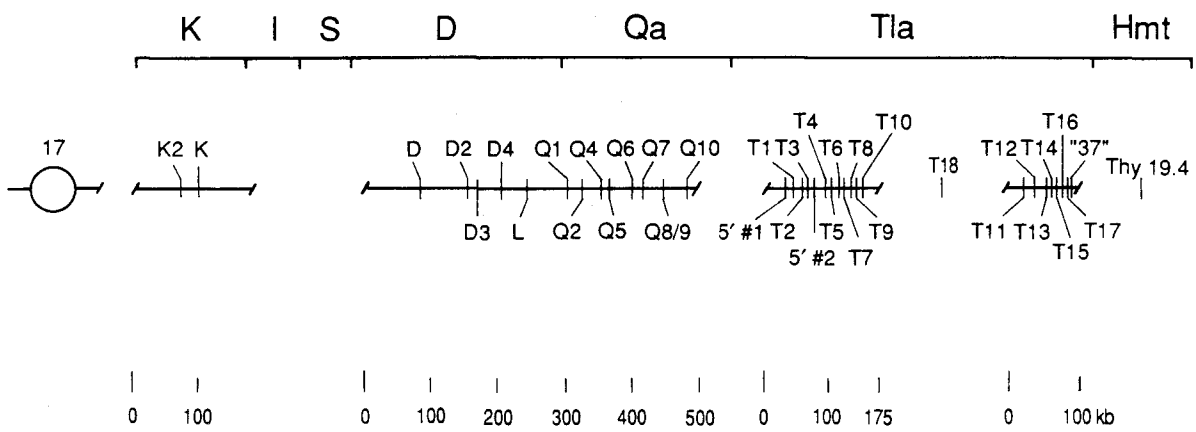

K
I
S
D
Qa
Tla
Hmt
17
K2 K
D D2 D4 Q1 Q4 Q6 Q7 Q10
D3 L Q2 Q5 Q8/9
T4 T10
T1 T3 T6 T8
5′ #1 T2 T5 T9
5′ #2 T7
T18
T16
T12 T14 "37"
Thy 19.4
T11 T13 T17
T15
0 100
0 100 200 300 400 500
0 100 175
0 100 kb

Figure 4. *Exon/intron organization of typical class I gene, H-2D^d.* Exons are indicated as boxes while non-coding sequences are shown as horizontal lines.


1
2
3
4
5
6
7
8
0
1000
2000
3000
4000bp

Chapter 1

Comparison of Exon 5 Sequences From 35 Class I Genes of the BALB/c Mouse

In press at *Journal of Experimental Medicine*

COMPARISON OF EXON 5 SEQUENCES FROM 35 CLASS I GENES OF THE BALB/c MOUSE

By KURT A. BRORSON, STEPHEN W. HUNT III**, TIM HUNKAPILLER, Y. HENRY SUN+, HILDE CHEROUTRE#, DEBORAH A. NICKERSON, AND LEROY HOOD

From the Division of Biology, 147-75, California Institute of Technology, Pasadena, California 91125, U.S.A.

**Present address: Division of Rheumatology and Immunology, The University of North Carolina, Chapel Hill, North Carolina 27599, U.S.A.

+Present address: Institute of Molecular Biology, Academia Sinica, Nankang, Taipei 11529, Taiwan, R.O.C.

#Present address: Department of Microbiology and Immunology, University of California, Los Angeles, California 90024, U.S.A.

The mouse class I major histocompatibility complex (MHC) molecules are structurally related 45 kilodalton (kD) cell-surface glycoproteins that associate non-covalently with β_2 microglobulin, a 12 kD polypeptide (1). Class I molecules can be divided into two groups based on their pattern of expression and their function. The transplantation antigens, H-2K, H-2D, and H-2L, are expressed on most somatic cells, and present viral antigens to cytotoxic T lymphocytes (2). The other group, the non-classical class I antigens, exhibit a generally more restricted tissue distribution and are probably not involved in antigen presentation (3-7).

The BALB/c mouse has at least 35 class I genes that map to five genetic loci: *K*, *D*, *Qa*, *Tla* and *Hmt* (8-11; Fig. 1). The classical transplantation antigen genes, K^d, D^d, and L^d map to the *K* and *D* loci, as do four other class I genes; $K2^d$, $D2^d$, $D3^d$ and $D4^d$ (12,13). The *Qa* and *Tla* loci together contain 28 known class I genes including some shown to encode non-classical class I antigens (6,14-16). In BALB/c mice, the eight *Qa* region genes are named $Q1^d$, $Q2^d$, $Q4^d$, $Q5^d$, $Q6^d$, $Q7^d$, $Q8/9^d$ and $Q10^d$, and the 19 *Tla* region genes are named $T1^c$ through $T18^c$ and 37^c. The newly described *Hmt* region contains at least three class I genes (10), including the *Thy19.4* gene (11), which is included in this study.

Class I genes contain 6 to 8 exons (14,17,18). Exon 1 encodes a hydrophobic leader segment, which is proposed to assist in the transport of the molecule to the cell surface and is cleaved posttranslationally (19). Exons 2, 3, and 4 each encode the three 90 amino acid external domains: $\alpha 1$, $\alpha 2$, and $\alpha 3$. A short external connecting peptide, as well as the transmembrane domain and part of the cytoplasmic segment that includes charged anchoring residues, is encoded by exon 5 (Fig. 2). Exons 6, 7 and 8 encode the remainder of the cytoplasmic domain. Analysis of exon 5 sequences shows that they are generally not conserved for direct sequence similarity, but rather for maintaining hydrophobicity in the

transmembrane stretch (20). Certain class I gene products, like those of the $Q4^d$ and $Q10^d$ genes, are secreted and do not maintain a hydrophobic transmembrane domain. The $Q7^d$ gene product, the Qa-2 antigen, has a typical hydrophobic transmembrane domain and charged anchor residues, yet is linked to the cell surface via a phosphatidylinositol linkage (15). The transmembrane domain of the $Q7^d$ molecule is proposed to be cleaved before expression on the cell surface.

This report compares the exon 5 DNA sequences of the 35 known class I genes of the BALB/c mouse. Such a comparison can reveal which of these exons can encode a hydrophobic transmembrane, and whether the putative gene product could be membrane-bound or secreted. Whereas the structure of the external α1, α2, and α3 domains has been resolved for at least one human class I antigen (21), no direct structural data exist for the transmembrane domains for the class I molecules. Therefore, an analysis of the predicted amino acid sequences of these genes could reveal what amino acid sequence and structural considerations are important for the function of the transmembrane domains. Analysis of the sequences reveals that in spite of extensive nucleotide sequence divergence, only four class I gene fifth exons, those from the $Q10^d$, $T5^c$, $T11^c$, and $T12^c$ genes, have frame shifts or stop codons that terminate their translation and prevent them from encoding a domain that is hydrophobic and long enough to span a lipid bilayer. Of the remaining fifth exons, 27 can encode membrane-spanning domains that resemble those of the classical transplantation antigens in that they can be divided into a proline-rich connecting peptide, a transmembrane segment, and a cytoplasmic segment with anchoring basic residues. In addition, hydrophobic moment analysis of the predicted transmembrane domains reveals that several, including those of the Qa-2 and TL antigens, are sufficiently amphipathic to promote intramembrane protein interactions. The conservation of the ability to

encode a potentially functional transmembrane domain in the majority of the fifth exons suggests that selective pressure exists on them to remain functional, possibly because the majority of class I genes, including the divergent ones, are functionally important.

Materials and Methods

Sequencing of Transmembrane Exons. Individual class I genes or gene fragments were cloned from BALB/c MHC class I cosmids (8) into M13mp18 or pUC18 derived vectors. DNA sequencing was performed using the dideoxynucleotide chain termination method (22). Sequencing was primed with an oligonucleotide (5' ACCTTCCAGAAGTGGGCA 3') derived from a conserved area of the fourth exon of the L^d gene (23). This primer was chosen because the same sequence occurs in the fourth exon of several divergent class I genes including the $H\text{-}2K^d$, D^d, L^d, $T13^c$, and $Q7^d$ (24) genes, and hence is presumably highly conserved in most class I genes. Since exon 5 is generally about 120 nucleotides long and 210 nucleotides downstream of the primer, it was possible to determine the complete exon 5 sequence of all unpublished genes on one strand with one set of sequencing reactions. Exon 5 sequences that could not be directly aligned with previously reported sequences were also sequenced on the opposite strand using complementary oligonucleotide primers derived from intron 5 sequences. The fifth exons that can be aligned and were not sequenced on two strands were those from the $K2^d$, $Q1^d$, $Q4^d$, $Q5^d$, $Q6^d$, $Q10^d$, and $T3^c$ genes.

Sequence Alignments and Comparisons. Sequence alignments were performed using the method of Needleman and Wunch (25), which inserts gaps into one or the other of the sequences in a pairwise comparison to maximize the similarity between the two sequences. In the percentage sequence similarity

calculation, a gap of any size is counted as one mismatch, whereas unmatched sequences at either end are not counted.

After alignment, pairs of sequences were analyzed at each position for possible and observed silent and replacement substitutions (26). A single base change that does not change the predicted translation of coding region sequence is considered to be silent, while one that does change the predicted translation is considered to be a replacement. Each possible pairing of aligned sequences was analyzed, and substitutions were totaled for each category.

Analysis of Translated Exon 5 Sequences. The translated exon 5 sequences were analyzed by an algorithm that calculates the hydrophobicity of 21 amino acid stretches of the sequence (27). The hydrophobicity values of individual amino acids are taken from a consensus scale adapted from five separate hydrophobicity measurements (28). The method calculates which 21 amino acid stretch has the highest hydrophobicity value, thereby predicting which segment, if any, best defines the transmembrane domain.

Within the predicted 21 amino acid transmembrane segment, the hydrophobic moments, a measure of amphipathicity, of 11 amino acid stretches were calculated using the equation of Eisenberg *et al.* (27). The highest hydrophobic moment value for each transmembrane was plotted against the hydrophobicity value for the corresponding 11 amino acid stretch. The empirically defined area of the graph in which the point falls predicts where the predicted helix is likely to be found relative to the membrane, and whether it resides in the membrane alone or in association with another protein.

Results and Discussion

Exon 5 Sequences and Groups. The DNA sequences of the fifth exons of 35

BALB/c class I genes are shown in Figure 3. In most cases, intron 4 and a portion of intron 5 are also included. The sequences were either obtained in this study from subclones of the BALB/c cosmids, or from previously published reports. In most cases, exon 5 is identified by nucleotide similarity to known fifth exons, while in the cases of the *T7*^{C}, *T15*^{C}, and *Thy19.4* genes, exon 5 is identified by the hydrophobicity of the translation, and the relative position 3' of exon 4. Exon boundaries are identified by comparison to class I genes for which spliced cDNA clones have been isolated (29,30, Hunt *et al.* in preparation), and by position of consensus splice sites. Donor splice sequences are not found in the *T4*^{C}, *T5*^{C}, *T7*^{C}, and *T15*^{C} fifth exons. The fifth exons of the *T7*^{C} and *T15*^{C} genes are interrupted by a B1 short interspersed repetitive element (31) after 143 base pairs (bp), while those of the *T4*^{C} and *T5*^{C} genes are similar to other fifth exons for the first 46 and 58 bp, respectively, but contain non-homologous sequences beyond what appears to be a recombinational or gene conversion boundary. The predicted reading frames are identified by the hydrophobicity of the translated amino acids, and by conformity to the reading frame established by the fourth exon.

The fifth exon sequences are assigned to the same group if they share at least 75% similarity with each other (32). The exon five sequences can be assigned to seven non-overlapping groups. The largest group includes all of the *H-2* and *Qa* loci genes, and in addition, several even-numbered *Tla* region genes: *T4*^{C}, *T6*^{C}, *T8*^{C}, *T10*^{C}, *T14*^{C}, *T16*^{C}, and *37*^{C}. A second group includes the *T1*^{C}, *T3*^{C}, *T11*^{C}, and *T13*^{C} genes , while a third includes the *T2*^{C}, *T5*^{C}, and *T12*^{C} genes. Finally, the *T9*^{C}/*T17*^{C} and *T7*^{C}/*T15*^{C} gene pairs form two additional groups, while the *T18*^{C} and *Thy19.4* genes form two additional single gene groups. Consensus sequences are derived for each group, based on the most frequent nucleotide used at each position.

Nucleotide sequence similarity among members of each group ranges from 73 to 99% (Table Ia). No two members of any group are exactly identical, making the exon 5 sequences diagnostic for the identification of BALB/c class I genes. Among members of each group, several types of mutational events have occurred subsequent to the duplications that created them, including nucleotide substitutions and short deletions. In addition, the exon 5 of the $Q1^d$ gene has an extra 18 bp that matches 15 of the 18 nucleotides immediately following it and thus it is probably the product of an internal sequence duplication. Interestingly, an 18 bp insertion that matches the $Q1^d$ insertion in 13 of 18 nucleotides is also found in the same position in a rat class I gene (33; J. Howard, personal communication). Since the inserted sequence is about as similar to the rat sequence as it is to the 18 bp sequence following it, it is possible that this duplication event occurred before mouse/rat divergence. Alternatively, since this appears to be a single mutational event in both species, it is possible that the duplications occurred independently. Since the insertion in the $Q1^d$ exon 5 does not match precisely either the sequences immediately following it, or the homologous insertion in the rat fifth exon, it is unclear which of these two possibilities is the case.

Comparison of group consensus sequences reveals that some groups are related while others appear not to be (Table Ib). Groups 2 and 3 are about 73% similar to each other. Their members are different enough from each other to be classified as distinct groups based on the criteria of this report, but they are clearly evolutionarily related. The other groups are possibly related to each other since some pairings share as much as 54% similarity. Unlike the similarity between groups 2 and 3, it is unclear if 32% to 54% sequence similarity between these groups is the result of divergent evolution from an ancestral exon 5, or

rather of convergent evolution of unrelated transmembrane exons. Codon usage that is restricted to maintain hydrophobicity of the translation could result in unrelated sequences attaining greater than random similarity. Thus, it is conceivable that the exon 5 sequences have multiple origins as the result of exon shuffling or *de novo* generation, and share similarity because of convergent evolution. This is most possible for the $T18^{C}$ fifth exon since it shares only 32% to 43% similarity to all of the other groups.

The existence of variation in the transmembrane exons in the class I gene family argues that their most important sequence consideration is the retention of a hydrophobic translation (20,34). Extensive variation occurs in transmembrane domains when their only function is to anchor a protein to a membrane. To test whether selective pressure exists for the fifth exon sequences to retain their translation, mutation frequencies of silent and replacement sites were determined for the group 1, 2, and 3 fifth exons (Table II). Since the members of groups 2 and 3 can be aligned, they were pooled to maximize the number of sites tested. As a contrast, silent and replacement site mutation frequencies were also determined in the exon 4 read-through frame of intron 4. If selective pressure is exerted on a coding region sequence, the frequency of mutation at silent sites is predicted to be higher than that of replacement sites. As expected, silent and replacement sites have approximately equivalent frequencies of mutation in the intron 4 sequences. However, in the fifth exons of groups 1, 2, and 3, replacement sites have a higher frequency of mutation than silent sites. This suggests that there is little selective pressure to maintain their protein encoding sequences other than for hydrophobicity, although the variation in the putative *Tla* region gene transmembranes may have evolved because they perform specialized functions that are different from those of the transplantation antigens.

In contrast to the fifth exon sequences, almost all exon 4 sequences are at least 80% similar to each other (Hunt *et al.* in preparation, K. Brorson, unpublished), supporting the concept that all of the class I genes evolved from a common ancestor (35). It is interesting that the highly conserved fourth exons and the highly divergent fifth exons are separated only by a 120 nucleotide intron. Dot-matrix identity plots between group consensus sequences (Fig. 4) reveal that between groups, intron 4 is more conserved than exon 5, and that there are two general areas of conservation. One area is the splice-acceptor site and the first approximately 10 bp of exon 5. The other area is the 5' portion of intron 4, adjacent to the conserved exon 4. It is conserved among all of the genes except the *T18*^{c} gene, and the *Thy19.4* gene where only the middle of the intron is conserved. Intron 4 is also more conserved than exon 5 when compared among groups (Table II). However, it is unlikely that this reflects selective pressure for their conservation, as would occur if read-through translation from exon 4 is important, since the synonymous substitution frequency of the read-through frame is approximately equal to the non-synonymous substitution frequency in group 1 as well as in groups 2 and 3. Instead, the distinct breaks in similarity evident in the dot-matrix identity plots suggest that the conservation in the fourth intron is a result of recombinational events involved in the evolution of the class I gene family. These recombination events could have included transmembrane exon shuffling or *de novo* generation events that created hybrid genes with similarity to classical class I genes in exon 4 and the 5' portion of intron 4, but little or no similarity in exon 5 and the rest of intron 4. Alternatively, it is suggested that short introns in class I genes are generally more conserved than the interior portions of long introns, because proposed recombination events that transfer exons between class I genes could often extend beyond the end of the exons into a

portion of the surrounding introns (36). It is conceivable that the 5' portions of the fourth introns are generally more conserved than the 3' portions because such DNA segment exchange events could occur more often between fourth exons than fifth exons (14).

The exon 5 sequence data can be used to support models for the evolution of specific groups of class I genes. It is proposed that the $Q4^b$-$Q10^b$ genes in the C57BL/10 mouse resulted from duplications of a primordial *Qa* gene pair, with the even- and odd-numbered genes derived from one or the other of the primordial genes (37,38). The *Q8* and *Q9* genes were subsequently fused in an unequal crossover event to form the hybrid gene $Q8/9^d$ of BALB/c mice (37). The exon 5 sequence data support this model since exon 5 in the $Q4^d$ and $Q6^d$ genes share 97% similarity and a single nucleotide deletion, causing a frame shift at nucleotide 62. In addition, the $Q7^d$ and $Q8/9^d$ genes are 99% similar to each other in exon 5. However, the $Q5^d$ gene is 98% similar to the D^d gene in exon 5, suggesting that it had undergone a DNA segment exchange event from that gene.

The two gene clusters, $T1^c$-$T10^c$ and $T11^c$-37^c (Fig. 1), of the *Tla* region of the BALB/c mouse, are proposed to have resulted from a duplication of an entire block of genes (14). The exon 5 sequence groups 2, 3, 4, and 5 define gene pairs with representatives in the same order on both clusters. These gene pairs are $T1^c/T11^c$, $T2^c/T12^c$, $T3^c/T13^c$, $T6^c/T14^c$, $T7^c/T15^c$, $T8^c/T16^c$, and $T9^c/T17^c$. The placement of these pairs in a specific order in both clusters supports the cluster duplication model. Because of similarities in restriction enzyme site patterns, the $T4^c$ and $T5^c$ genes are proposed to have been created in a duplication of a pair of genes that also produced the $T6^c$ and $T7^c$ genes (14). However, the exon 5 sequence data do not support this contention since the $T5^c$ exon 5 does not resemble the $T7^c$ exon 5, but instead is 95% similar to the $T2^c$

exon 5 in the first 58 nucleotides. Beyond that point, it does not resemble any other exon 5 sequence. In addition, exon 5 in the $T4^C$ gene is 93% similar to that of the $T6^C$ gene in the first 46 nucleotides, after which is a 21 bp polythymidine tract followed by a non-homologous sequence. Since both the $T4^C$ and $T5^C$ exon 5 sequences are interrupted by non-homologous sequences, it is likely that instead of being created as a block duplication of the $T6^C$ and $T7^C$ genes, the $T4^C$ and $T5^C$ genes are partial class I genes that were duplicated separately from distinct sources. The $T5^C$ gene is probably a partially duplicated $T2^C$ or $T12^C$ gene, while the $T4^C$ gene is probably a partially duplicated group 1 class I gene.

Splice Junction Sequences. The putative acceptor-splice junction sequences of 35 and donor sequences of 30 of the class I fifth exons are shown in Figure 5. The fifth exons of the $T4^C$, $T5^C$, $T7^C$, and $T15^C$ genes do not have donor-splice sequences, probably because they were eliminated by recombination or repetitive element integration events during their evolution. In addition, since *Thy19.4* transcripts do not splice exon 5 to any 3' exons (11), it is also excluded from the donor-sequence figure. All 35 acceptor sequences have polypyrimidine tracts of 16 to 58 bp in length, followed by an AG dinucleotide. Since splicing invariably occurs following an AG dinucleotide in eukaryotic genes (39), and polypyrimidine tracts in acceptor-splice signals are generally over 11 bp in length, all 35 acceptor-splice sequences appear functional. Similarly, all of the donor splice sequences match the consensus sequence ($^{C}_{A}$AG/GT$^{A}_{G}$AGT) with at least 6 of 9 nucleotides and have the invariant GT dinucleotide at the immediate splice junction. Since donor-splice sequences need to match the consensus sequence in only 5 of the 9 nucleotides to be functional (40), and since all of the class I donor sequences have the invariant GT dinucleotide found in all eukaryotic donor sequences (39), all 30 of the class I donor sequences also appear to be functional.

Since none of the splice sequences in the 35 class I fifth exons appear abnormal, it is unlikely that any of the fifth exons will be non-functional because of splicing abnormalities.

In addition to donor- and acceptor-splice sequences homologous to those in previously characterized genes, several possible alternative splice sites can be identified (Fig. 3). These sites include in-frame donor sequences in members of groups 1-5 in intron 4 near the junction with exon 4. If a class I transcript splices at these sites, between 2 and 5 amino acids would be added to the α3 domain of its translation product. In the $T2^c$, $T5^c$, $T12^c$, and $T18^c$ genes there are additional possible alternative donor sites in intron 4 that generate a different frame from that identified in cDNA transcripts. However, transcripts that splice these possible alternative donor sites to possible alternative acceptor sites in the 5' portion of exon 5 would place the hydrophobic translation of exon 5 in the same reading frame as exon 4. The translation would be slightly longer in exon 4 and slightly shorter in exon 5. The translation of the $T12^c$ fifth exon terminates two amino acids after the acceptor-splice signal homologous to those characterized in other class I genes. However, a $T12^c$ transcript could encode a hydrophobic transmembrane if spliced at the possible alternative splice sites. Finally, there are donor signals in the fifth exon of the $T4^c$ gene. However, since these sites were probably introduced to this gene by a recombination event, it is unclear if they actually evolved to splice the $T4^c$ fifth exon to a 3' exon.

Analysis of Predicted Protein Sequence. Translation of the DNA sequences reveals that, with the exception of the $Q10^d$, $T5^c$, $T11^c$, and $T12^c$ genes, each class I gene has an open reading frame in exon 5 whose translation is potentially hydrophobic and long enough to span a lipid bilayer (27; Fig. 6). Thus, each of these 31 fifth exon-encoded amino acid sequences can be divided into connecting,

transmembrane and cytoplasmic segments. In this study the transmembrane segment is arbitrarily defined as the most hydrophobic 21 amino acids of the fifth exon translation, since that is the chain length required to form an α-helix that can completely span a lipid bilayer (41). Analysis of previously characterized membrane-bound proteins reveals that transmembrane domains range in length from 20 to 28 amino acids (42). Since the choice of 21 amino acids is arbitrary, it is important to note that it is possible that in the actual gene products, some of the residues near the calculated borders may not reside in their predicted environment. The exceptions to the arbitrary assignment of 21 amino acids are the predicted $Q1^d$, $T7^c$, and $T15^c$ transmembrane domains, which clearly have hydrophobic segments in excess of 27 amino acids. In the translated sequences, the putative connecting peptides vary from 4 to 20 amino acids in length. The putative cytoplasmic portions are between 4 and 13 amino acids in length, except in the translations of the $Q4^d$, $Q6^d$, $T4^c$ and $T15^c$ fifth exons where none can be identified.

Connecting Peptides. Analysis of the putative connecting peptides reveals that the amino acid usage is similar to that in the hinge regions of immunoglobulins (24; Table III). Proline (28%) is the most commonly used amino acid. In addition, asparagine (8%) is also present in these segments. These amino acids tend to disrupt any helical structure that may form in the junction between the transmembrane and outer domains (43). In addition, serine and threonine predominate in the connecting peptide at 16% and 14%, respectively. These two amino acids with small polar hydroxyl side chains are common in exposed areas, and their presence is not predicted to contribute to or disrupt the formation of α helices (43). However, comparison of 31 proteins exhibiting segment flexibility demonstrates that both serines and threonines tend to be concentrated in flexible

segments (44). The serines and threonines in the connecting peptides could confer more flexibility to this segment.

The imposition of a flexible β-turn structure in the connecting peptide could facilitate stretching and pivoting at this segment in a manner similar to that in the hinge region of immunoglobulins. It is suggested that the freedom of movement of the two immunoglobulin Fab arms relative to the Fc stem results from proline-rich amino acid sequences within the hinge segment that favor flexibility (24,45,46). This freedom of movement of the Fab arms is believed to be important for the function of immunoglobulins (47). Similarly, in the case of transplantation antigens, freedom of movement in the connecting peptide could be important to facilitate interaction with the T-cell receptor. It is interesting that the $T18^C$ molecule is predicted to have a connecting peptide 20 amino acids in length. It would be twice as long as those in the transplantation antigens, but it is unclear if there is any significance to this difference.

Transmembrane Segments. In the predicted transmembrane segments, four hydrophobic amino acids dominate: valine (26%), alanine (15%), isoleucine (14%) and leucine (13%) (Table III). Phenylalanine is present at an intermediate level of 8%, but the other three hydrophobic amino acids, tryptophan, methionine, and proline, each constitute 4% or less of the transmembrane amino acids. Proline (1%) may be absent from the transmembrane segments because its structure may tend to disrupt the α helical structure assumed by the hydrophobic amino acids in their aliphatic environment (48,49) although it has been suggested, given the work with bacterial transmembrane domain deletion mutants, that transmembrane domains are not always completely α-helical (50). Tryptophan (3%) and methionine (4%) are used less often in proteins in general, and their lower usage in the transmembrane domains could reflect this (51). In addition, tryptophan has

been suggested to be more hydrophilic than previously believed, and is represented infrequently in other transmembrane segments (52). Glycine (8%) is the only non-hydrophobic amino acid found in abundance in the transmembrane segments. Glycine has a very small slightly polar side chain which would probably not significantly decrease the hydrophobicity of the transmembrane segments. Interestingly, the putative $Q1^d$, $T7^c$, and $T15^c$ molecules are predicted to have hydrophobic transmembrane segments of between 27 and 49 amino acids. Although these genes have not been shown to encode proteins, if they did, it would be interesting to see how such long hydrophobic segments are accommodated in the membrane.

Proteins with transmembrane domains that interact with other proteins within the lipid bilayer, as class II MHC molecules are proposed to (53), are thought to do so because they contain short stretches within their membrane-spanning segment that are sufficiently amphipathic to promote such interactions (27). To test to see whether any of the class I transmembrane segments as well as four BALB/c class II transmembrane segments, A^d_α, A^d_β, E^d_α, and E^d_β, could interact within the membrane with other proteins, the hydrophobic moment, a measure of amphipathicity, was calculated for 11 amino acid stretches within them. The length of 11 amino acids corresponds to approximately 3 turns of an α helix, which is believed to be the typical amphipathic segment size that interacts non-covalently with other proteins. For each transmembrane segment, the 11 amino acid stretch with the highest hydrophobic moment was determined, and that hydrophobic moment value was plotted against the hydrophobicity value for the 11 amino acid stretch (Fig. 7). Whether an 11 amino acid stretch is predicted to be sufficiently amphipathic to promote interactions within the membrane depends on which empirically defined area within the graph its plotted

point falls (27).

This analysis reveals that the plotted points of all four class II transmembranes fall within or near the area defined as multimeric transmembrane. Since class II molecules are dimeric on the cell surface, it is proposed that the transmembrane's amphipathicity and amino acid sequence conservation is consistent with the hypothesis that they are dimeric within the lipid bilayer as well (53). On the other hand, the algorithm predicts that several class I transmembrane domains are not sufficiently amphipathic to be predicted to interact with other proteins within the membrane. The transplantation antigens, K^d, D^d and L^d, are heterodimers with β_2-microglobulin, a small polypeptide with no membrane-spanning segment. Therefore, the prediction that they do not have amphipathic transmembrane segments is consistent with their probable monomericity within the membrane. In addition to the transplantation antigens, the putative molecules encoded by the majority of group 1 genes and the *T9*^{c} and *T17*^{c} genes are also predicted to be monomeric within the membrane. However, other putative class I transmembrane segments are predicted to be sufficiently amphipathic to associate within the membrane with other proteins. These include those predicted to be encoded by the *Q4*^{d}, *Q7*^{d}, *Q8/9*^{d}, *T3*^{c}, *T7*^{c}, *T10*^{c}, *T13*^{c}, *T15*^{c}, *T18*^{c}, and *Thy19.4* genes. The *Q4*^{d} gene encodes a secreted class I product, and the *Q7*^{d} gene encodes the Qa-2 antigen, which is linked to the cell surface by a phosphatidylinositol linkage. The amphipathicity data are consistent with the hypothesis that during the processing or transport of these two molecules, intramembrane interactions occur with yet uncharacterized proteins. In addition, the putative products of several *Tla* region class I genes, including the *T13*^{c} gene that encodes the TL^c antigen, are also predicted to interact with other proteins within the membrane. These *Tla* region genes, *T3*^{c}, *T7*^{c}, *T13*^{c}, *T15*^{c}, *T18*^{c}, and

Thy19.4, also have highly divergent fifth exons, suggesting that the putative molecules encoded by these genes perform functions different from those of the classical class I molecules. The hydrophobic moment data are consistent with the hypothesis that their putative function may require intramembrane interactions with other molecules, possibly for the initiation of signaling cascades.

Cytoplasmic Segment. In the cytoplasmic portion, two basic amino acids predominate: arginine (25%) and lysine (19%). Basic amino acids are commonly found on the cytoplasmic side of transmembrane domains and are proposed to prevent the short cytoplasmic domain from being pulled through the hydrophobic lipid bilayer (42,54). Histidine, a slightly basic amino acid is not present in the class I anchor sequences. It is probably too weakly basic to serve in an anchor sequence. Also present in the cytoplasmic segments are methionine (11%), valine (8%) and asparagine (11%). It is unclear if there is any significance to the presence of these amino acids, although it is interesting that methionines and asparagines are clustered at both ends, but not at the center, of the transmembrane domains. Some of these residues may be spacer at the end of the domain and between the highly charged basic residues. Others are transmembrane segment amino acids included in cytoplasmic portion because of the arbitrary decision to limit the transmembrane segment to the 21 most hydrophobic amino acids.

Four of the 31 potential hydrophobic transmembrane domains do not have basic anchoring residues at the carboxyl-terminal end: $Q4^d$, $Q6^d$, $T4^c$, and $T15^c$. It is known that the $Q4^d$ gene, like the $Q10^d$ gene, encodes a secreted class I molecule (7,55). The lack of an anchor sequence probably contributes to the fact that it is a secreted class I, in spite of its hydrophobic transmembrane segment. On the other hand, studies with $H\text{-}2L^d$ gene mutants demonstrate that anchoring

residues are not absolutely necessary for cell-surface expression of class I glycoproteins (56). In addition, it is suggested that the $Q4^P$ product can exist on the cell surface of transfected cells (57). Clearly, the absence of anchoring residues can not be used universally as criteria for whether a class I molecule is secreted or membrane-expressed. If there are products of the $Q6^d$, $T4^c$, and $T15^c$ genes, it will be interesting to see whether they are secreted, membrane-bound, or both. Interestingly, the transmembrane domains of the putative $T7^C$, $T9^C$, and $T17^C$ molecules are predicted to end with only one or two basic anchoring residues, whereas the classical transplantation antigens are anchored by three to four basic residues (Fig. 6). If these genes can encode class I molecules, it will also be interesting to see if these molecules are anchored to the cell membrane as efficiently as the transplantation antigens. The $Q10^d$ product has neither a hydrophobic transmembrane nor anchoring residues (55), and it is known to be a secreted product (6).

Implication of Predicted Protein Sequences. This analysis reveals that almost all of the 35 class I genes have fifth exons that have open reading frames that could potentially encode a domain that is sufficiently hydrophobic and long to span a lipid bilayer, and hence by this criterion appear to be functional. Only four of the 35 BALB/c class I gene fifth exons, those of the $Q10^d$, $T5^c$, $T11^c$, and $T12^c$ genes, appear to be exceptions and have stop codons or frame shifts that prevent them from encoding a hydrophobic transmembrane domain. However, this does not necessarily imply that these four genes are pseudogenes, since at least the $Q10^d$ gene encodes a presumably functional soluble class I molecule. Thus, given the analysis of these sequences, it is not evident that any of these genes are pseudogenes. Of the remaining 31 class I genes, 27 have a fifth exon that could encode a domain with the transplantation antigen motif of maintaining both hinge-

like connecting peptides and basic anchor amino acids at the appropriate ends of the hydrophobic stretch. Only the putative $Q4^d$, $Q6^d$, $T4^c$, and $T15^c$ transmembrane domains are exceptions by lacking basic anchor amino acids. Overall, the amino acid usage of these segments is appropriate for their predicted function. The hingelike segments use amino acids expected to introduce β turns and segmental flexibility. The transmembrane segments consist of hydrophobic amino acids, while there are anchoring basic residues in the cytoplasmic segments. The maintenance of this motif is particularly striking because of the extensive sequence divergence of the fifth exon groups. Since the majority of the fifth exons appear to be able to encode a functional transmembrane domain, it is unlikely that their divergence is merely a result of genetic drift in the absence of selective pressure. It is more likely that selective pressure exists to maintain their functionality, suggesting that the majority of the class I genes, including the divergent ones, are functionally important. It could be speculated that the fifth exons have diverged from each other because the molecules that they encode have specialized functions other than antigen presentation to T cells. If the molecules that the divergent groups encode are involved in other functions, the fifth exons would still be selected for the ability to encode a transmembrane domain, but not one similar to those in restriction elements. Clearly, analysis of exon 5 alone can only suggest what a particular class I gene can encode; further sequence and expression studies will be required to determine the extent of expression of class I genes.

Summary

DNA sequences of the fifth exon, which encodes the transmembrane domain, were determined for the BALB/c mouse class I MHC genes and used to study the

relationships between them. On the basis of nucleotide sequence similarity, the exon 5 sequences can be divided into seven groups. Although most members within each group are at least 80% similar to each other, comparison between groups reveals that the groups share little similarity. However, in spite of the extensive variation of the fifth exon sequences, analysis of their predicted amino acid translations reveals that only four class I gene fifth exons have frameshifts or stop codons that terminate their translation and prevent them from encoding a domain that is both hydrophobic and long enough to span a lipid bilayer. Exactly 27 of the remaining fifth exons could encode a domain similar to those of the transplantation antigens in that it consists of a proline-rich connecting peptide, a transmembrane segment, and a cytoplasmic portion with membrane-anchoring basic residues. The conservation of this motif in the majority of the fifth exon translations in spite of extensive variation suggests that selective pressure exists for these exons to maintain their ability to encode a functional transmembrane domain, raising the possibility that many of the non-classical class I genes encode functionally important products.

The authors wish to thank Drs. I. Stroynowski, M. Zuniga, K. Fischer Lindahl, and J. Kobori for critically reviewing this manuscript, Dr. J. Howard for critical insights on exon 5 evolution and for sharing unpublished rat exon 5 sequence data, and Mrs. C. Blagg for expert secretarial assistance. D. Nickerson was a visiting associate from the University of South Florida, Tampa, Florida 33620, U.S.A.

References

1. Silver, J., and L. Hood. 1974. Detergent solubilized H-2 alloantigen is associated with a small molecular weight polypeptide. *Nature (Lond.)* **249**:764.

2. Zinkernagel, R. M., and P. C. Doherty. 1980. MHC-restricted cytotoxic T cells: studies on the role of polymorphic major transplantation antigens determining T-cell restriction specificity, function, and responsiveness. *Adv. Immunol.* **27**:51.

3. Old, L. J., E. A. Boyse, and E. Stockert. 1963. Antigenic properties of experimental leukemias. I. Serological studies *in vitro* with spontaneous and radiation-induced leukemias. *J. Natl. Cancer Inst.* **31**:977.

4. Flaherty, L. 1976. The *Tla* region of the mouse: Identification of a new serologically defined locus, *Qa-2*. *Immunogenetics* **3**:533.

5. Stanton, T. H. and E. A. Boyse. 1976. A new serologically defined locus, *Qa-1*, in the *Tla* region of the mouse. *Immunogenetics* **3**:525.

6. Maloy, W., J. 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.

7. Robinson, P. J. 1985. Qb-1, a new class I polypeptide encoded by the *Qa* region of the mouse *H-2* complex. *Immunogenetics* **22**:285.

8. Steinmetz, M., A. Winoto, K. Minard, and L. Hood. 1982. Clusters of genes encoding mouse transplantation antigens. *Cell* **28**:489.

9. Winoto, A., M. Steinmetz, and L. Hood. 1983. Genetic mapping in the major histocompatibility complex by restriction enzyme polymorphism: most mouse class I genes map to the *Tla* complex. *Proc. Natl. Acad. Sci. USA* **80**:3425.

10. Richards, C. S., M. Bucan, K. Brorson, M. Kiefer, S. Hunt, H. Lehrach, and K. Fischer Lindahl. Genetic and molecular mapping of the *Hmt* region of the mouse. *EMBO J*, in press.

11. Brorson, K., S. Richards, S. Hunt, H. Cheroutre, K. Fischer Lindahl, and L. Hood. Analysis of a new class I gene mapping to the *Hmt* region of the mouse. *Immunogenetics*, in press.

12. Goodenow, R., M. McMillan, M. Nicolson, B. 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.

13. Stephan, D., H. Sun, K. Fischer Lindahl, E. Meyer, G. Hämmerling, L. Hood, and M. Steinmetz. 1986. Organization and evolution of *D* region class I genes in the mouse major histocompatibility complex. *J. Exp. Med.* **163**:1227.

14. Fisher, D. A., S. W. Hunt, and L. 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.

15. Stroynowski, I., M. Soloski, M. Low, and L. Hood. 1987. A single gene encodes soluble and membrane-bound forms of the major histocompatibility Qa-2 antigen: Anchoring of the product by a phospholipid tail. *Cell* **50**:759.

16. Robinson, P., D. Bever, A. Mellor, and E. Weiss. 1988. Sequence of the mouse *Q4* class I gene and characterization of the gene product. *Immunogenetics* **27**:79.

17. Steinmetz, M., K. W. Moore, J. Frelinger, B. T. Sher, F. 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, 1981.

18. Kvist, S., L. Roberts, and B. Dobberstein. 1983. Mouse histocompatibility genes: structure and organization of a K^d gene. *EMBO J.* **2**:245.

19. Blobel, G., and B. Dobberstein. 1975. Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *J. Cell Biol.* **67**:835.

20. Uehara, H., J. Coligan, and S. Nathenson. 1981. Isolation and sequence analysis of the intramembranous hydrophobic segment of the H-$2K^b$ murine histocompatibility antigen. *Biochemistry* **20**:5936.

21. Bjorkman, P., M. Saper, B. Samraoui, W. Bennett, J. Strominger, and D. Wiley. 1987. Structure of the human class I histocompatibility antigen, HLA-A2. *Nature (Lond.)* **329**:506.

22. Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe, 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol.* **143**:161.

23. 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., D.C.)* **215**:679.

24. Kabat, E., T. Wu, M. Reid-Miller, H. Perry, and K. Gottesman. 1977. *Sequences of Proteins of Immunological Interest*, 4th ed. U.S. Department of Health and Human Services, Public Health Services, National Institutes of Health.

25. Needleman, S., and C. Wunsch. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**:443.

26. Kimura, M. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. *Proc. Natl. Acad. Sci. USA* **78**:454.

27. Eisenberg, D., E. Schwarz, M. Komaromy, and R. Wall. 1984. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol.* **179**:125.

28. Eisenberg, D., R. Weiss, T. Terwilliger, and W. Wilcox. 1982. Hydrophobic moments and protein structure. *Faraday Symp. Chem. Soc.* **17**:109.

29. Steinmetz, M., J. Frelinger, D. Fisher, T. Hunkapiller, D. Pereira, S. Weissman, H. Uehara, S. Nathenson, and L. Hood. 1981. Three cDNA clones encoding mouse transplantation antigens: Homology to immunoglobulin genes. *Cell* **24**:125.

30. Chen, Y.-T., Y. Obata, E. Stockert, and L. 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.

31. Krayev, A., D. Kramerov, K. Skryabin, A. Ryskov, A. Bayev, and G. Georgiev. 1980. The nucleotide sequence of the ubiquitous repetitive DNA sequence B1 complementary to the most abundant class of mouse fold-back RNA. *Nucl. Acids Res.* **8**:1201.

32. Crews, S., J. Griffin, J. Huang, K. Calame, and L. Hood. 1981. A single V_H gene segment encodes the immune response to phosphoryl choline: Somatic mutation is correlated with the class of antibody. *Cell* **25**:59.

33. Kastern, W. 1985. Characterization of two class I major histocompatibility rat cDNA clones, one of which contains a premature termination codon. *Gene* **34**:227.

34. Davis, N., and P. Model. 1985. An artificial anchor domain: Hydrophobicity suffices to stop transfer. *Cell* **41**:607.

35. Klein, J. and F. Figueroa. 1986. The evolution of class I MHC genes. *Immunol. Today* **7**:41.

36. Hayashida, H., and T. Miyata. 1983. Unusual evolutionary conservation and frequent DNA segment exchange in class I genes of the major histocompatibility complex. *Proc. Natl. Acad. Sci. USA* **80**:2671.

37. 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 and evolution of the class I gene family in the major histocompatibility complex of the C57BL/10 mouse. *Nature (Lond.)* **310**:650.

38. Devlin, J., E. Weiss, M. Paulson, and R. Flavell. 1985. Duplicated gene pairs and alleles of class I genes in the *Qa* region of the murine major histocompatibility complex: a comparison. *EMBO J.* **4**:3203.

39. Breathnach, R., and P. Chambon. 1981. Organization and expression of eukaryotic split genes coding for proteins. *Ann. Rev. Biochem.* **50**:349.

40. Mount, S. 1982. A catalogue of splice junction sequences. *Nucl. Acids Res.* **10**:459.

41. Tanford, C. 1980. *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*, Wiley, New York.

42. Warren, G. 1981. Membrane proteins: Structure and assembly. *In Membrane Structure* (Finean, J. B. and Michell, R. H., eds). Elsevier/North Holland Biomedical Press, Amsterdam, Holland, pp. 215-257.

43. Chou, P., and G. Fasman. 1978. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol.* **47**:45.

44. Karplus, P., and G. Schulz. 1985. Prediction of chain flexibility in proteins. *Naturwissenschaften* **72**:212.

45. Seegan, G., C. Smith, and V. Schumaker. 1979. Changes in quaternary structure of IgG upon reduction of the interheavy-chain disulfide bond. *Proc. Natl. Acad. Sci. USA* **76**:907.

46. Marquart, M., J. Deisenhofer, R. Huber, and W. Palm. 1980. Crystallographic refinement and atomic models of the intact immunoglobulin molecule Kol and its antigen-binding fragment at 3.0Å and 1.9Å resolution. *J. Mol. Biol.* **141**:369.

47. Klein, M., N. Haeffner-Cavaillon, D. Isenman, C. Rivat, M. Navia, P. Davies, and K. Dorrington. 1981. Expression of biological effector functions by immunoglobulin G molecules lacking the hinge region. *Proc. Natl. Acad. Sci. USA* **78**:524.

48. Henderson, R., and P. N. T. Unwin. 1975. Three-dimensional model of purple membrane obtained by electron microscopy. *Nature* **257**:28.

49. Guidotti, G. 1977. The structure of intrinsic membrane proteins. *J. Supermolecular Structure* **7**:489.

50. Davis, N., J. Boeke, and P. Model. 1985. Fine structure of a membrane anchor domain. *J. Mol. Biol.* **181**:111.

51. Klapper, M. 1977. The independent distribution of amino acid near neighbor pairs into polypeptides. *Biochem. Biophys. Res. Commun.* **78**:1018.

52. Clothia, C. 1976. The nature of the accessible and buried surfaces in proteins. *J. Mol. Biol.* **105**:1.

53. Malissen, M., T. Hunkapiller, and L. Hood. 1983. Nucleotide sequence of a light chain gene of the mouse *I-A* subregion: A_{β}^{d}. *Science (Wash. D.C.)* **221**:750.

54. Tomita, M., and V. Marchesi. 1975. Amino acid sequence and oligosaccharide attachment sites of human erythrocyte glycophorin. *Proc. Natl. Acad. Sci. USA* **72**:2964.

55. 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.

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

57. Schepart, B., J. Woodward, M. Palmer, M. Macchi, P. Basta, E. McLaughlin-Taylor, and J. Frelinger. 1985. Expression in L cells of transfected class I genes from the mouse major histocompatibility complex. *Proc. Natl. Acad. Sci. USA* **82**:5505.

58. Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts, and J. Watson. 1983. *Molecular Biology of the Cell.* Garland Publishing. New York, New York.

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

60. Transy, C., S. R. Nash, B. David-Watine, M. Cochet, S. W. Hunt, L. E. Hood, and P. Kourilsky. 1987. A low polymorphic mouse *H-2* class I gene from the *Tla* complex is expressed in a broad variety of cell types. *J. Exp. Med.* **166**:341.

61. Headly, M., S. Hunt, K. Brorson, J. Andris, L. Hood, J. Forman, and P. Tucker. 1989. DNA sequence analysis of $D2^d$: A new *D*-region class I gene. *Immunogenetics* **29**:359.

62. Fisher, D. A., M. Pecht, M., and L. Hood. 1989. DNA sequence of a class I pseudogene from the *Tla* region of the murine MHC: Recombination at a B2 Alu repetitive sequence. *J. Mol. Evol.* **28**:306.

Table Ia

Percent similarity of fifth exons within groups.

Group 1

K2	86																					
D	82	84																				
D2	82	85	82																			
D3	85	85	85	87																		
D4	87	90	85	85	85																	
L	82	85	85	85	88	85																
Q1	85	85	85	85	91	87	85															
Q2	89	89	86	88	91	90	90	91														
Q4	84	86	78	85	87	86	88	85	88													
Q5	83	84	98	83	86	85	85	85	87	79												
Q6	83	84	78	84	86	86	87	83	87	97	78											
Q7	83	86	82	88	90	86	87	88	90	87	83	86										
Q8/9	83	85	81	87	89	86	86	87	84	86	82	87	99									
Q10	73	78	76	82	79	79	84	78	82	80	77	79	81	80								
T4	89	89	78	80	87	87	85	89	93	85	78	83	83	80	76							
T6	81	84	81	81	83	86	83	88	88	81	82	80	81	80	75	93						
T8	86	87	86	83	86	86	86	85	89	81	86	81	83	82	79	87	85					
T10	83	85	82	85	85	83	83	87	89	80	83	79	84	83	76	85	81	82				
T14	82	85	82	82	84	87	85	88	89	84	83	83	82	81	76	93	95	86	82			
T16	87	88	87	84	87	87	87	86	90	82	87	81	84	83	80	89	86	99	83	87		
37	88	88	87	86	92	89	87	92	94	89	88	88	88	87	79	93	90	89	88	91	90	
	K	K2	D	D2	D3	D4	L	Q1	Q2	Q4	Q5	Q6	Q7	Q8/9	Q10	T4	T6	T8	T10	T14	T16	

Group 2

T3	90			
T11	94	93		
T13	91	98	95	
	T1	T3	T11	

Group 3

T5	95	
T12	98	91
	T2	T5

Group 4

T7 vs. T15: 98

Group 5

T9 vs. T17: 99

In comparisons, only the first 58 bp of the $T5^c$ fifth exon and 46 bp of the $T4^c$ fifth exon are used. The full lengths of the fifth exons are compared between all of the other genes.

Table Ib

Percent similarity between consensus sequences of fifth exon groups.

Group 2	54					
Group 3	53	73				
Group 4	50	43	43			
Group 5	41	38	48	46		
Group 6	32	41	33	36	43	
Group 7	48	38	49	40	42	37
	G1	G2	G3	G4	G5	G6

Prior to the percent similarity calculation, the fifth exon sequences were aligned with gaps to maximize the result. Group 1 through 7 consensus sequences are abbreviated as G1 through G7.

Table II
Replacement and silent mutation frequencies in exon 5 and intron 4.

	Replacement (Rep)	Silent (Sil)	Possible Rep/Sil	Observed Rep/Sil
Exon 5 Group 1	2766/17,249 16%	765/6133 12%	2.81	3.62
Exon 5 Groups 2 & 3 (aligned)	202/1259 16%	40/361 11%	3.49	5.05
Intron 4 (Exon 4 read-through) Group 1	1264/15,319 8%	594/5951 10%	2.57	2.13
Intron 4 (Exon 4 read-through) Groups 2 & 3 (aligned)	217/1,568 14%	68/496 14%	3.16	3.19

Frequencies are expressed both as a fraction of observed changes over possible changes, and as a percentage. Ratios of possible and observed replacement and silent changes are also shown. The $T4^C$ and $T5^C$ fifth exons were omitted from this analysis since their 3' portions were created by recombination events, not duplication and point mutation.

Table III
Amino Acid Compositions of Predicted Connecting, Transmembrane, and Cytoplasmic Segments

	Connecting Segment	Transmembrane	Cytoplasmic Portion
Acidic			
Aspartic Acid (D)	2.5%	0.3%	0%
Glutamic Acid (E)	0%	0.4%	1.0%
Basic			
Lysine (K)	1.4%	0.1%	19.0%
Arginine (R)	2.8%	0.6%	25.1%
Histidine (H)	0.7%	0%	1.0%
Polar			
Glycine (G)	0.7%	8.3%	3.1%
Asparagine (N)	8.2%	0.9%	10.8%
Glutamine (Q)	3.2%	0.1%	0%
Cysteine (C)	0.4%	1.6%	1.0%
Serine (S)	16.0%	1.7%	3.1%
Threonine (T)	14.2%	1.6%	6.7%
Tyrosine (Y)	2.8%	0.1%	0.5%
Non-Polar			
Alanine (A)	2.8%	15.1%	1.0%
Valine (V)	4.6%	26.5%	7.7%
Leucine (L)	2.8%	12.9%	5.6%
Isoleucine (I)	0%	14.2%	2.1%
Proline (P)	28.4%	1.3%	0%
Phenylalanine (F)	0.7%	7.8%	0%
Methionine (M)	7.1%	3.5%	11.3%
Tryptophan (W)	0.7%	2.8%	1.0%

The calculation reflects percent representation in a total of 31 connecting and transmembrane segments, and 27 cytoplasmic segments. The individual amino

acids have been previously assigned to acidic, basic, polar, or non-polar categories (58).

Figure 1. Map of class I genes in the BALB/c MHC. Class I genes map to the *K*, *D*, *Qa*, *Tla* and *Hmt* regions. The *I* and *S* regions contain class II and complement genes, respectively. The order of the *Tla* region gene clusters is unknown, as is the distance between the *K*, *D*, *Tla* and *Hmt* regions. The upper line represents the genetic map and the gene clusters are indicated below.

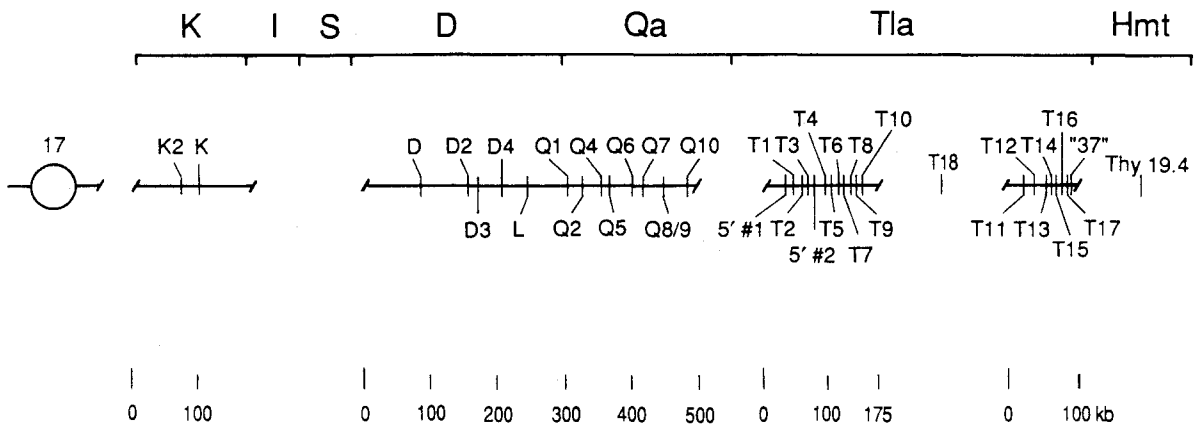

K
I
S
D
Qa
Tla
Hmt
17
K2 K
D D2 D4 Q1 Q4 Q6 Q7 Q10
D3 L Q2 Q5 Q8/9
T4 T10
T1 T3 T6 T8
5′ #1 T2 T5 T9
5′ #2 T7
T18
T16
T12 T14 "37"
T11 T13 T17
T15
Thy 19.4
0 100
0 100 200 300 400 500
0 100 175
0 100 kb

Figure 2. Model of typical class I transmembrane domain: H-2K^{d}. Connecting, transmembrane, and cytoplasmic segments are shown in which environment they are predicted to reside by the criteria used in this report (27).

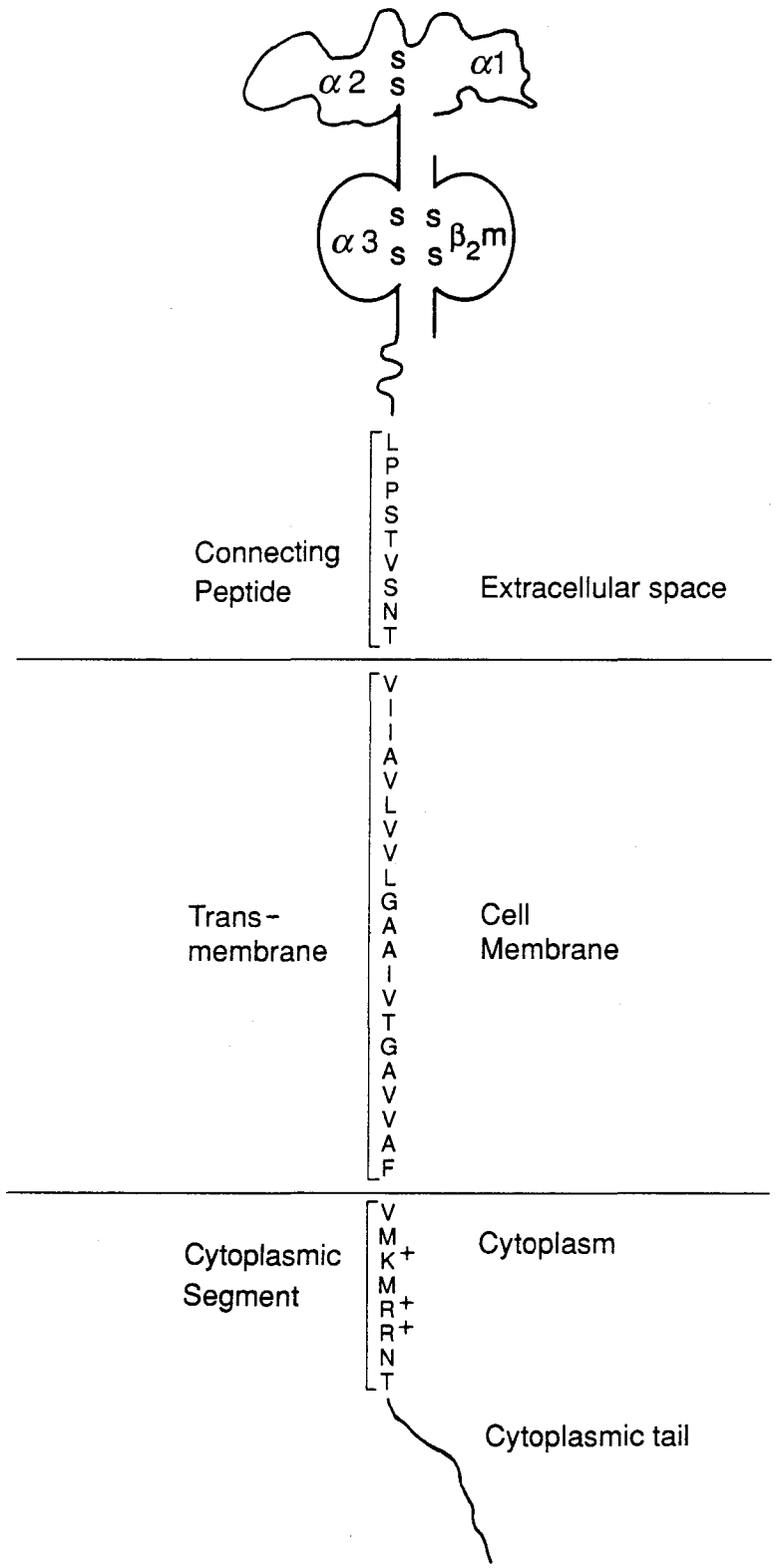
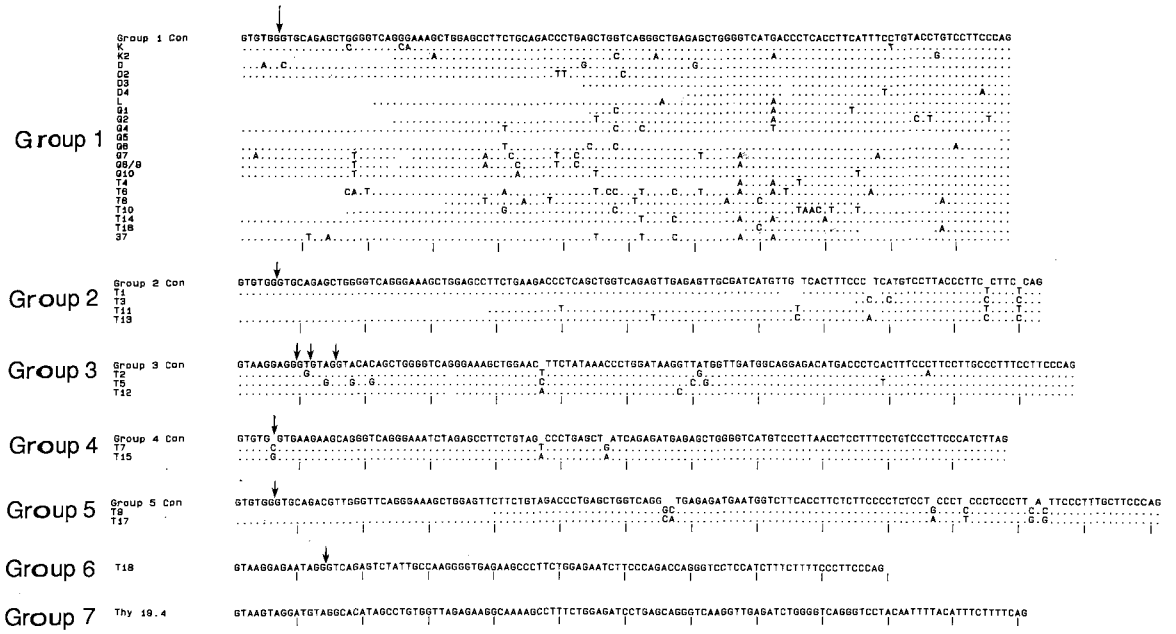
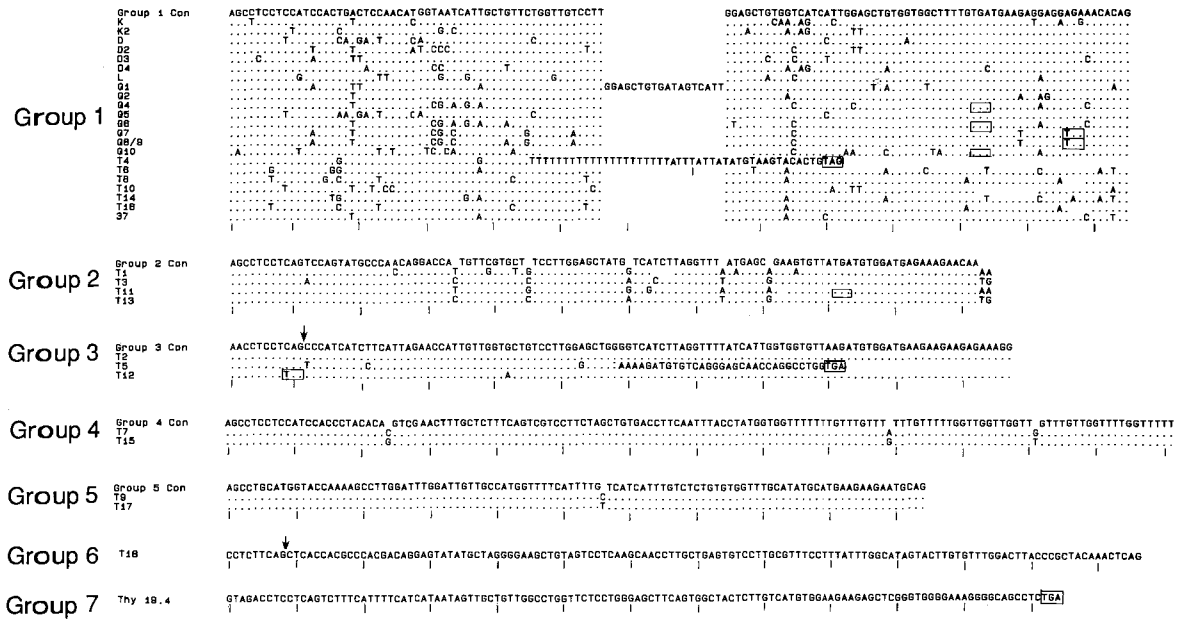

α2
S
S
α1
α3
S
S
S
S
β2m
L
P
P
S
T
V
S
N
T
Connecting
Peptide
Extracellular space
V
I
I
A
V
L
V
V
L
G
A
A
I
V
T
G
A
V
V
A
F
Trans-
membrane
Cell
Membrane
V
M
K+
M
R+
R+
N
T
Cytoplasmic
Segment
Cytoplasm
Cytoplasmic tail

Figure 3. The DNA sequence of fifth exons of the 35 BALB/c class I genes. Genes that were sequenced on one strand are the $K2^d$, $Q1^d$, $Q4^d$, $Q5^d$, $Q10^d$, and $T3^c$ genes. All others were sequenced on both strands or have been previously published (L^d, (22); K^d, (17); $T13^c$, (13); D^d, (59); 37^c, (60); *Thy19.4*, (11); $T1^c$, (62); $D2^d$, (61); $Q7^d$, $T9^c$, $T17^c$, $T18^c$, $D3^d$, $D4^d$, Hunt *et al.* in preparation; $Q5^d$, $Q6^d$, $Q8/9^d$, I. Stroynowski, personal communication; $Q10^d$, N. Ulker, personal communication). Intron 4 and 5 sequences are included in most cases. In frame translation termination codons found in the fifth exons of the $Q4^d$, $Q6^d$, $Q7^d$, $Q8/9^d$, $Q10^d$, $T4^c$, $T5^c$, $T7^c$, $T11^c$, $T12^c$, $T15^c$, and *Thy19.4* genes are boxed. In the cases of the $T4^c$, $T5^c$, $T7^c$, and $T15^c$ genes, the separation between exon 5 and intron 5 is arbitrary. In the $T4^c$ and $T5^c$ genes, the end of exon 5 is defined as the inframe stop codon where translation terminates, while in the $T7^c$ and $T15^c$ genes, the B1 repeat element is arbitrarily included with the other intron 5 sequences. Possible alternative splice signals are indicated by arrows above the consensus sequences.

INTRON 4


Group 1
Group 2
Group 3
Group 4
Group 5
Group 6
Group 7
Group 1 Con
Group 2 Con
Group 3 Con
Group 4 Con
Group 5 Con
T18
Thy 19.4

EXON 5



INTRON 5

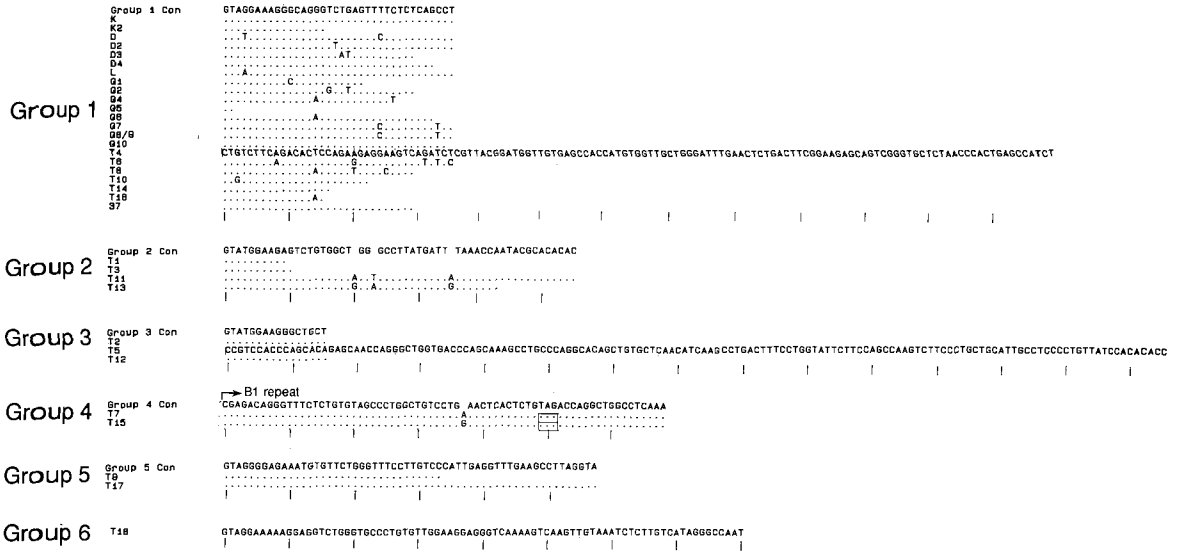


Figure 4. Dot-matrix identity comparison of group 1 consensus sequences with groups 2-7 consensus sequences. Exon 5 boundaries are shown on the top and on the side. Each dot represents a 6 of 8 nucleotide match between the sequences.

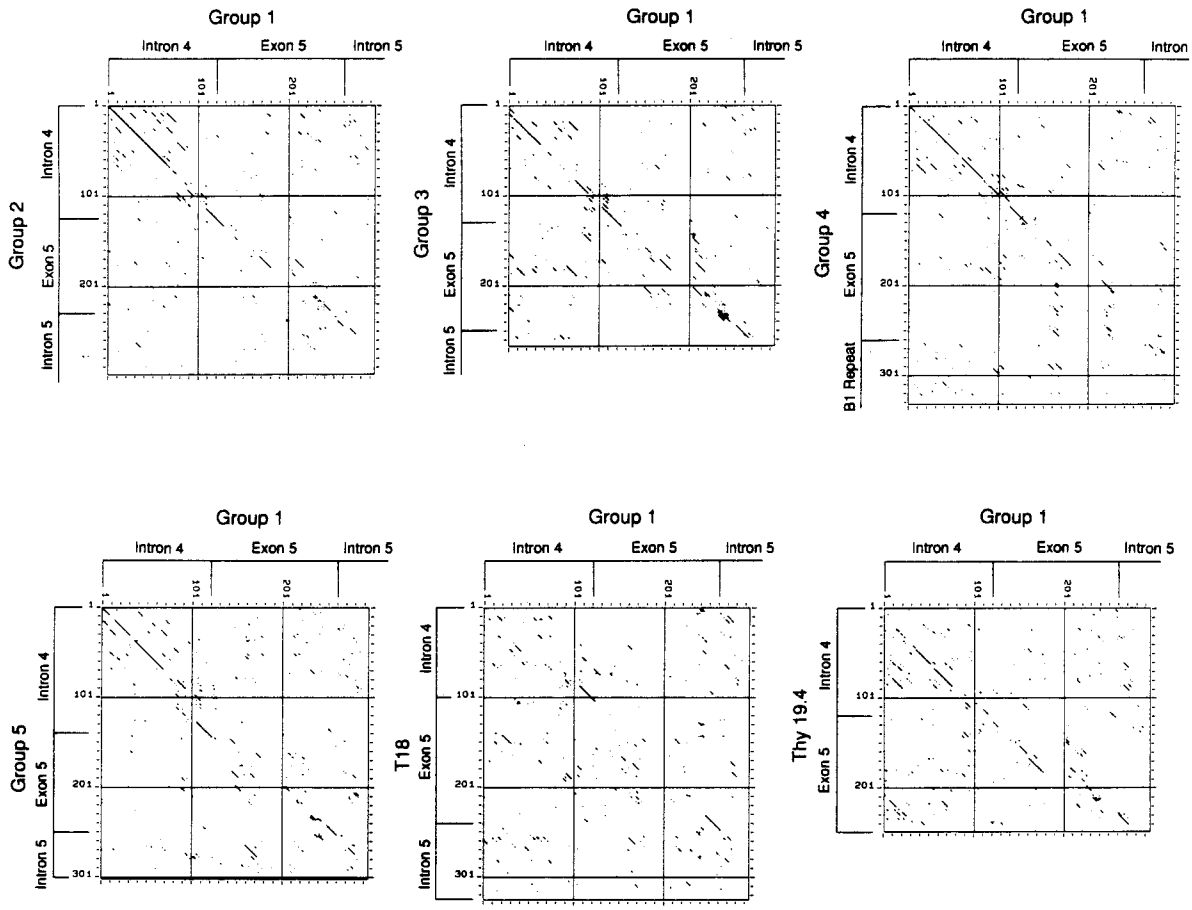

Group 1
Intron 4
Exon 5
Intron 5
Group 2
Group 3
Group 4
B1 Repeat
Group 5
T18
Thy 19.4
101
201
301

Figure 5. The acceptor and donor splice sequences of BALB/c class I genes. The consensus sequence appears above the compiled sequences, and vertical lines indicate where splicing is predicted to occur.


Acceptor Sequences
Donor Sequences
Acceptor Consensus
Donor Consensus
K
K2
D
D2
D3
D4
L
Q1
Q2
Q4
Q5
Q6
Q7
Q8/9
Q10
T4
T6
T8
T10
T14
T16
37
T1
T3
T11
T13
T2
T5
T12
T7
T15
T9
T17
T18
Thy 19.4

Figure 6. The translations of the exon 5 sequences. Where they exist, borders between predicted connecting, transmembrane, and cytoplasmic segments are indicated by vertical lines between amino acids. Frameshifts present in the fifth exons of the $Q4^d$, $Q6^d$, and $Q10^d$ genes cause a portion of their exon 5 translation to be non-homologous to those of other class I genes in their group, while recombination events that occurred in the $T4^c$ and $T5^c$ fifth exons produced a similar result for their translations.

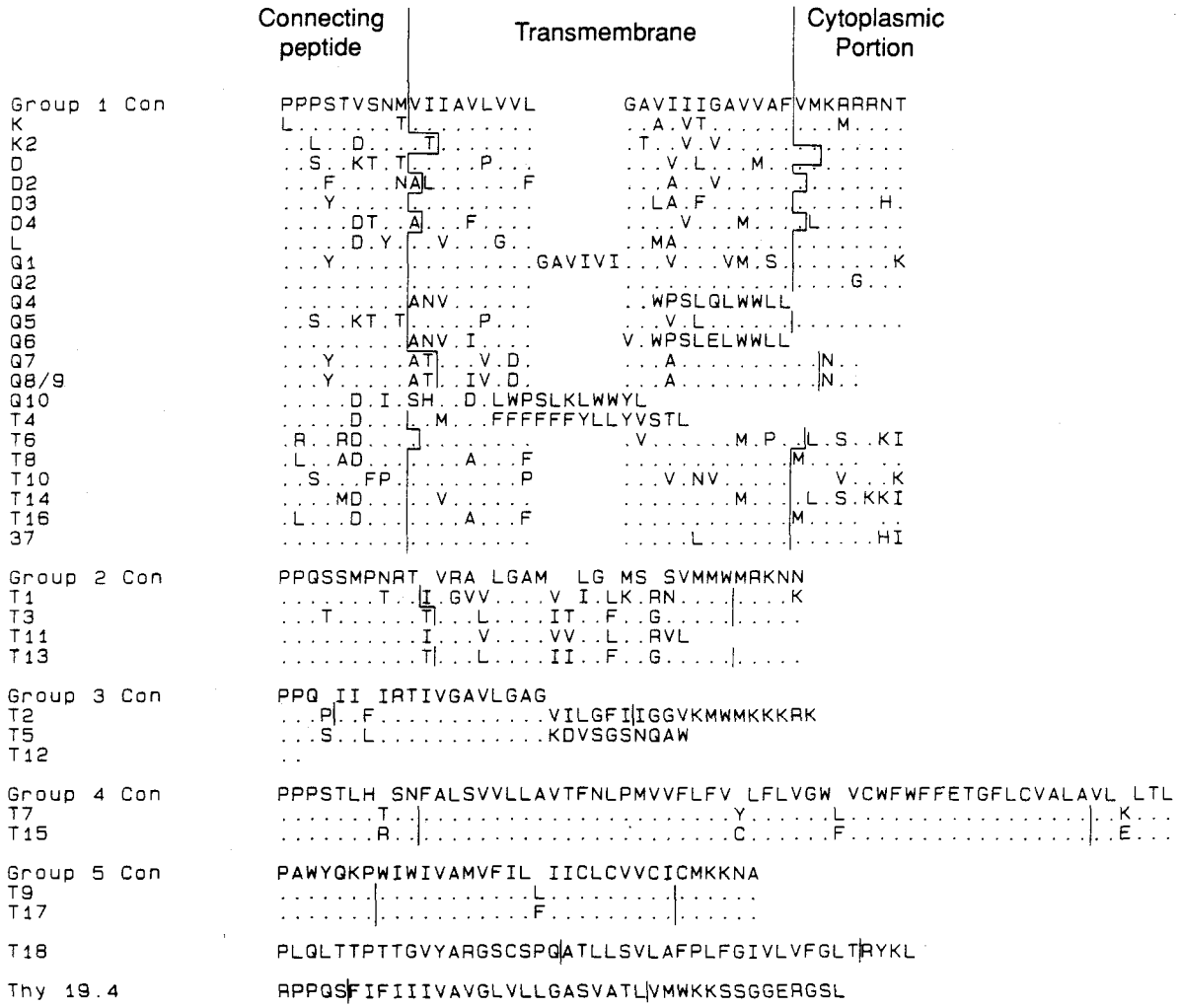
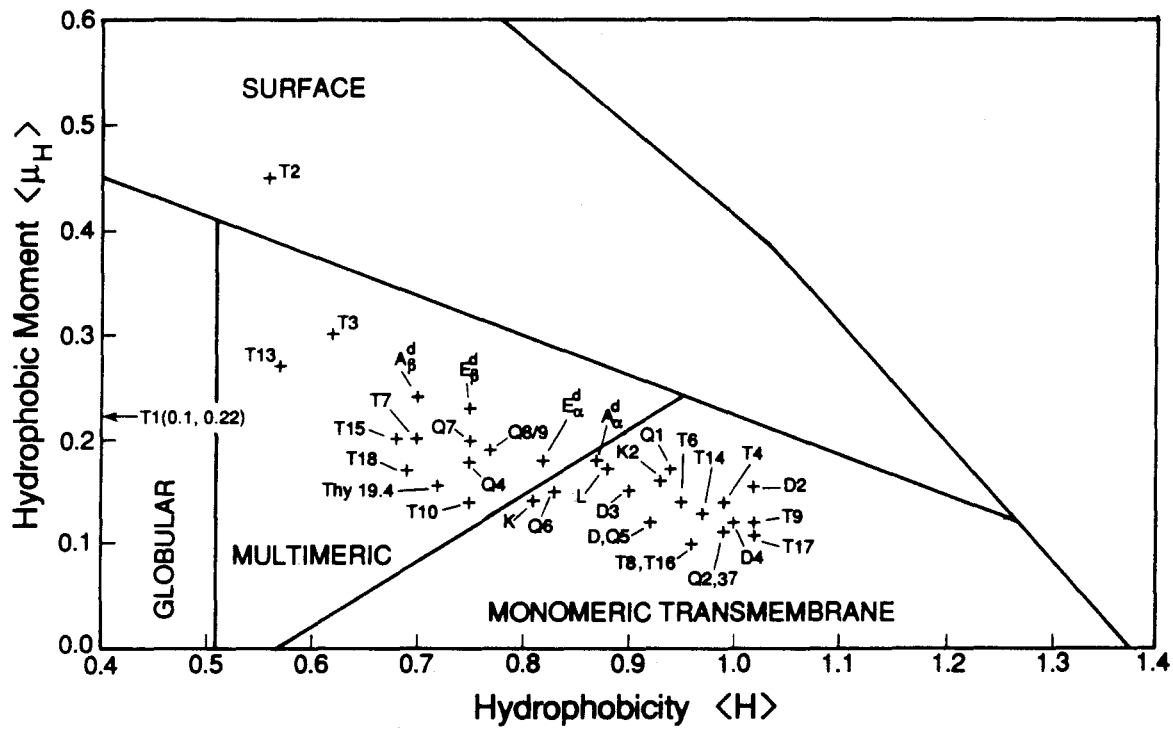

Connecting peptide
Transmembrane
Cytoplasmic Portion
Group 1 Con
K
K2
D
D2
D3
D4
L
Q1
Q2
Q4
Q5
Q6
Q7
Q8/9
Q10
T4
T6
T8
T10
T14
T16
37
Group 2 Con
T1
T3
T11
T13
Group 3 Con
T2
T5
T12
Group 4 Con
T7
T15
Group 5 Con
T9
T17
T18
Thy 19.4

Figure 7. Hydrophobic moment plot. Hydrophobic moment $\langle \mu_H \rangle$ is plotted against hydrophobicity <H> for 11 amino acid segments within each of 31 class I fifth exon translations, as well as the transmembrane domains of the class II molecules A^d_α, A^d_β, E^d_α, and E^d_β. Each point plots within arbitrary areas labeled: surface, globular, multimeric or monomeric transmembrane. Although arbitrarily defined by Eisenberg *et al.* (27), 36 of 49 transmembrane segments originally used to define these regions were correctly plotted within the region of the graph that corresponded to their type.


SURFACE
MULTIMERIC
GLOBULAR
MONOMERIC TRANSMEMBRANE
T1(0.1, 0.22)
T2
T3
T13
T7
T15
T18
Thy 19.4
T10
Q7
Q4
Q8/9
K
Q6
L
D3
D,Q5
T8,T16
Q2,37
D4
T17
T9
D2
T4
T14
T6
Q1
K2
Hydrophobic Moment <μH>
Hydrophobicity <H>
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
1.3
1.4

Chapter 2

Expression of Class I Major Histocompatibility Complex Genes in the BALB/C Mouse: At Least 15 Genes are Transcribed in the Adult Thymus

In preparation for *The Journal of Experimental Medicine*

EXPRESSION OF CLASS I MAJOR HISTOCOMPATIBILITY COMPLEX GENES IN THE BALB/C MOUSE: AT LEAST 15 GENES ARE TRANSCRIBED IN THE ADULT THYMUS**

BY STEPHEN W. HUNT III*‡, KURT A. BRORSON*, HILDE M. CHEROUTRE*#, AKIHIRO MATSUURA+, FUNG-WEN SHEN+, AND LEROY E. HOOD*

*From the *Division of Biology, 147-75, California Institute of Technology, Pasadena, California and +Department of Molecular Genetics, Showa University Research Institute, St. Petersburgh, Florida 33761*

‡Present address: Department of Medicine, The University of North Carolina, Chapel Hill, North Carolina 27599

#Present address: Department of Microbiology and Immunology, The University of California, Los Angeles, California 90024

**Supported by NIH grant AI-17565. S.W.H. is a special fellow of the Leukemia Society of America.

INTRODUCTION

Murine class I major histocompatibility complex (MHC) molecules are heterodimeric glycoproteins consisting of a 45 kilodalton (kD) heavy chain associated non-covalently with a 12 kD polypeptide, β_2-microglobulin (1). Two types of class I molecules have been identified. The highly polymorphic classical transplantation antigens, H-2K, D and L, are integral membrane proteins expressed on most somatic cells (2) that present viral and tumor antigens to cytotoxic T lymphocytes (3). In addition to the transplantation antigens, several structurally similar molecules have been identified that are less polymorphic and apparently are not viral antigen-presenting structures (66). They are the non-classical class I molecules that include the cell-surface Qa-1, Qa-2 and TL antigens (4-6), and the Qb-1 (7) and Q10 (8) molecules, which are secreted proteins. Although their class I-like structure suggests that some of these molecules are cell-surface recognition structures, their exact function has not been identified.

The genes encoding the class I molecules reside in the major histocompatibility complex on mouse chromosome 17 (63). The class I gene family has been studied extensively in BALB/c (9) and C57BL/10 (10) mice and contains 25-35 individual members. In the BALB/c mouse, class I genes map to the *K*, *D*, *Qa*, *Tla* and *Hmt* subregions of the MHC (11-14), with the majority mapping to the *Qa* and *Tla* subregions (Fig. 1). The classical transplantation antigen genes, *H-2K*^{d}, *D*^{d}, and *L*^{d}, map to the *K* and *D* loci (62), as do four other class I genes, *K2*^{d}, *D2*^{d}, *D3*^{d}, and *D4*^{d}. The *Qa* and *Tla* loci together contain 27 known class I genes, and include some shown to encode non-classical class I molecules (12,15). In BALB/c mice, the eight *Qa* region genes are the *Q1*^{d}, *Q2*^{d}, *Q4*^{d}, *Q5*^{d}, *Q6*^{d}, *Q7*^{d}, *Q8/9*^{d}, and *Q10*^{d} genes, while 19 *Tla* region genes are named the *T1*^{c} through *T18*^{c}

and 37^c. The *Hmt* subregion is a newly identified subregion containing at least three genes, including the *Thy19.4* gene described in this report (16).

Class I genes contain 6 to 8 exons (12,17). Exon 1 encodes a short hydrophobic leader segment that presumably assists in the transport of the molecule to the cell surface, and is cleaved posttranslationally (18). Exons 2, 3, and 4 each encode the three 90 amino acid external regions: α1, α2, and α3. The α3 region interacts most closely with β_2-microglobulin, and therefore has a highly conserved structure (19). An approximately 10 amino acid external connecting peptide, as well as the approximately 24 amino acid transmembrane region and part of the cytoplasmic segment that includes positively charged anchoring residues are encoded by exon 5. Exons 6, 7, and 8 encode the remainder of the approximately 40 amino acid cytoplasmic region.

The relatively large size of the class I gene family has prompted an analysis of the expression of these genes and a search for the proteins that they encode. These studies have relied mostly on the few serological reagents available, or DNA sequencing of class I genes. A few gene products or pseudogenes have been identified (Table I), but the majority of the 35 BALB/c class I genes have not been characterized extensively. One approach that has been used to determine which class I genes are expressed in a particular tissue is the analysis of class I transcripts from cDNA libraries. This approach identified six class I genes that are expressed in the liver of DBA/2 mice (20).

This report uses a similar approach to study class I gene expression in the thymus of the BALB/c mouse. Several non-classical class I molecules, including the Qa1, Qa2, and TL antigens, are expressed in the thymus, and have been proposed to be recognition structures involved in cell-cell recognition events. The thymus is the location of T-cell differentiation, and contains T cells at various

points of the differentiation process as well as other hematopoietic cells such as epithelial cells, dendritic cells, and macrophages (21-23). Hence, if non-classical class I molecules are involved in cell-cell recognition events, then the thymus would be a logical place to analyze class I gene expression because of the multiplicity of cells present there. This report presents evidence for the expression of at least 15 class I genes in the murine thymus, seven of which have not been previously characterized. The majority of these genes appear to have open reading frames, suggesting that they could encode class I molecules. In addition, many of them appear to be evolutionarily ancient, and hence, may have a primordial function common to many vertebrates.

MATERIALS AND METHODS

Enzymes and Reagents. Restriction and modifying enzymes were purchased from New England BioLabs (Beverly, MA), Boehringer Mannheim Biochemicals (Indianapolis, IN) or Bethesda Research Labs (Gaithersburg, MD). AMV reverse transcriptase was obtained from Life Sciences, Inc. (St. Petersburg, FL). Nitrocellulose membranes were purchased from Schleicher and Schuell (Keene, NH). Labeled α-[^{32}P]dCTP and γ-[^{32}P]dATP were obtained from Amersham (Arlington Heights, IL) and ICN (Irvine, CA), respectively. All other reagents were molecular biology grade.

Production of and Screening of cDNA Library. Thymus tissue was dissected from 5-6 week old BALB/cJ mice and rapidly frozen in liquid nitrogen. RNA was isolated by ethanol precipitation in 4 M guanidinium isothiocyanate (25), and polyadenylated RNA was isolated using HyBond-mAP paper (Amersham, Arlington Heights, IL). A λgt10 cDNA library was made from the thymus RNA using standard technologies (26). The cDNA library was screened using a 439 nucleotide

genomic fragment containing the fourth exon and portions of introns 3 and 4 of the BALB/c *H-2D*^{d} gene (27).

Identification of cDNA Clones. Plaque-purified clones were amplified and aliquots of the phage supernatant were filtered using a dot blot manifold (Schleicher and Schuell, Keene, NH) onto nitrocellulose . Phage supernatants were lysed and denatured *in situ* and the membranes were baked in a vacuum oven for 2 h at 80°C. Dot blots were hybridized with oligonucleotide probes derived from the fifth exon nucleotide sequence of the 35 known BALB/c class I genes (24). The oligonucleotide probes were labeled by phosphorylation by polynucleotide kinase (28) and used for hybridization at approximately 10^5 cpm/ml hybridization solution. Filters were washed in 2X SSC, 0.1% SDS at 50°C.

DNA Sequencing. Phage inserts were subcloned into M13mp18, M13mp19 or pGEM-1 (Promega-Biotec, Madison, WI) for sequence analysis by the dideoxynucleotide chain termination method (30). The DNA sequence of each cDNA clone presented in this report was obtained on two strands. A oligonucleotide derived from a highly conserved region of exon 4 having the sequence 5' ACCTTCCAGAAGTGGGCA 3' was used as a primer for DNA sequence analysis of exon 5.

Construction of Phylogenetic Trees. Synonymous replacement (nucleotide substitutions that do not alter the translation) frequencies were determined among pairings of genes in exons 2, 3 and 4 (59). These frequencies were corrected using a binomial equation $d = 3/4 \times \ln(1-4p/3)$ (47), and compiled into distance matrices. Rooted phylogenetic trees of exon 2 and 3 as well as exon 4 evolution were constructed from these distance matrices using the unweighted pair-group with arithmetic mean (UPGMA) method (48). Time points for branch nodes that follow the mammalian radiation 65 to 85 million years ago (42,43,45) were

assigned based on the previously reported mammalian synonymous mutation rate of 4.6×10^{-9} substitutions per site per year (49), while those that predate the mammalian radiation are not assigned a divergence date. Bird/mammal divergence is believed to have occurred 300 million years ago, on the basis of paleontological evidence (42).

RESULTS AND DISCUSSION

Isolation and Identification of class I cDNA Clones. A cDNA library of 3×10^5 clones was made with RNA isolated from 5-week-old BALB/c thymuses. This library was screened with a genomic fragment containing the fourth exon of the BALB/c *H-2D*^{d} gene (27). Since exon 4 is highly conserved, this probe is expected to hybridize to almost all class I genes. A total of 72 class I clones were isolated. These clones were initially sorted by hybridization to nineteen oligonucleotide probes derived from exon 5 sequences of individual BALB/c class I genes (data not shown, 24). In addition, several transcripts were identified by determining the DNA sequence of the fifth exon of the cDNA clone and comparing it to the previously reported genomic sequences (24). Transcripts identified by exon 5 sequencing were those from the *H-2D*^{d}, *D2*^{d}, *D3*^{d}, *D4*^{d}, *Q4*^{d}, *Q7*^{d}, *Q8/9*^{d}, *T9*^{c}, *T13*^{c}, *T17*^{c}, *T18*^{c}, *37*^{c}, and *Thy19.4* genes.

Range of Transcription of Class I Genes. Transcripts from 15 different class I genes were identified in the thymus library (Table II), and hence at least that many class I genes are expressed in the adult thymus. Approximately one-half of the transcripts are from the genes that encode the classical transplantation antigens K^d, D^d, and L^d. In addition, transcripts from 12 non-classical class I genes were isolated. The non-classical class I genes that are transcribed in the thymus are from the *D*, *Qa*, *Tla* and *Hmt* subregions of the murine MHC and

include the *T13*^{c} gene, which encodes the TL^c antigen (12), the *Q7*^{d} gene, which encodes the Qa-2 antigen (15), and the *Q4*^{d} gene, which encodes the secreted Qb-1 molecule (31). In addition, several transcripts from previously uncharacterized class I genes were isolated. To characterize the novel non-classical class I genes transcribed in the thymus, the DNA sequence of representative clones of the *D2*^{d}, *D3*^{d}, *D4*^{d}, *Q8/9*^{d}, *T9*^{c}, *T17*^{c}, *T18*^{c} and *Thy19.4* genes was obtained (Figure 2). In addition, the DNA sequence of a clone derived from the Qa-2 antigen gene *Q7*^{d} was obtained.

Within the sequences, exons 2, 3, and 4 are identified by nucleotide sequence similarity to those exons in the classical class I genes. The fifth exons are identified by comparison to previously reported exon 5 sequences (24). In most cases sequences matching the compiled intron 5 sequences are not present in the clone (24) and hence the sequences following the fifth exon are presumably exon 6. Several transcripts contain introns or are truncated at either the 5' or 3' end. In many cases, transcripts from the same gene were found in other libraries in a completely spliced form, and hence it is likely that transcripts from these genes normally splice out most of these introns. A summary of which exons and introns are present in the nine cDNA clones, and which sequences are putative introns is shown in Table III.

Comparison of the cDNA transcripts to the *H-2D*^{d} gene reveals that the non-classical class I genes belong to two types, on the basis of their nucleotide sequence similarity to the transplantation antigen genes (Table IVa). The transcripts from genes of the *H-2* and *Qa* subregions share significant nucleotide sequence similarity with the *H-2D*^{d} gene in each exon. These genes include the *D2*^{d}, *D3*^{d}, and *D4*^{d} genes from the *D* region, as well as the *Q4*^{d}, *Q7*^{d}, and *Q8/9*^{d} genes from the *Qa* region. With the exception of exon 7 of the *D2*^{d} gene, all of

the exons sequenced from these genes have open reading frames. However, the stop codon present in exon 7 would truncate the putative $D2^d$ molecule by only nine amino acids, and hence it is possible that the *$D2^d$* gene as well as the other genes could encode class I proteins.

A second type of non-classical class I gene expressed in the thymus is typified by the genes in the *Tla* and *Hmt* subregions of the MHC. These genes, *$T9^c$*, *$T13^c$*, *$T17^c$*, *$T18^c$*, and *Thy19.4* have significantly diverged from the classical class I genes and each other in nucleotide sequence, and thus, possibly in function as well (Table IVa). These genes share high (>80%) nucleotide sequence similarity with the *H-$2D^d$* gene in exon 4, and intermediate (50-70%) similarity in exons 2 and 3. The remaining portions of these genes are completely dissimilar to the *H-$2D^d$* gene, and to each other, and thus an alignment and direct comparison cannot be made. The only exceptions are the *$T9^c$* and *$T17^c$* genes, which are believed to have been created by a duplication event (12), and share greater than 94% nucleotide sequence similarity with each other (Table IVb). In spite of their divergence, all but one of these genes contain open reading frames as far as examined.

The *$T9^c$* gene, identified as a pseudogene in this study, has a frame shift in exon 3 which leads to a premature termination of translation. All of the exons in the other genes, with the exception of exon 7 in the *$D2^d$* gene, appear to contain open reading frames on the basis of partial DNA sequencing. The *$T17^c$* cDNA clone isolated from the thymus library does not contain the third exon, and hence it is not possible to exclude the possibility that this member of the *$T9^c$/$T17^c$* gene pair is also a pseudogene. However, sequencing of the genomic *$T17^c$* gene reveals that it has the same single-base deletion in exon 3 as the *$T9^c$* gene, suggesting that it too is a pseudogene (data not shown). In addition to a frame shift in exon

3, an apparent retrovirus integration event occurred in the first exon of the *T9*^{c} gene. The integrated retrovirus sequence present in the cDNA sequence is 89 bp long and is 78% identical to another endogenous retrovirus sequence found in the *Tla* region of C57BL mice (32). The integrated retroviral sequence also precludes the *T9*^{c} gene from encoding a protein since it introduces a stop codon into the homologous reading frame of exon 1.

Examination of aligned sequences reveals that several of the class I sequences contain insertions and deletions relative to the sequence of the *H-2D*^{d} gene. The majority of these insertions and deletions are multiples of three nucleotides, and thus would not alter the reading frame. The *T9*^{c} and *T18*^{c} genes have several insertions and deletions in their α1 and α2 domain-encoding second and third exons. However, in spite of five deletion and insertion events, as well as extensive sequence divergence, the *T18*^{c} gene has preserved an open reading frame, strongly suggesting that there is selective pressure to maintain its ability to encode a class I molecule.

Analysis of Translations of cDNA Clones. The aligned amino acid sequences of the putative molecules encoded by the cDNA clones, as well as portions of the human HLA-A2 (33) and BALB/c mouse H-2K^d, D^d, and TL^c (12,27,34) class I molecules are shown in Figure 3. Although it is a pseudogene, the translation of the *T9*^{c} gene is included since it is conceivable that it has a functional homolog in other mammals. 5' sequences that apparently correspond to at least part of the first exon of the *T18*^{c} gene are included in the *T18*^{c} cDNA clone. These sequences have an open reading frame in the same frame as exon 2, and include an initiation ATG codon. However, the translation of the putative exon 1 is neither long enough nor hydrophobic enough to serve as a leader sequence (18), and hence it is unclear whether these sequences correspond to the entire leader sequence, or

only to a portion of it.

Interestingly, in spite of extensive amino acid sequence divergence, the putative translation products of several of the *Tla* region genes generally appear to conserve several residues in the α1 and α2 domains that are critical in the formation of the tertiary structure of the human HLA-A2 class I molecule (19, P. Bjorkman, personal communication). These residues include positions in the HLA-A2 molecule at turns such as prolines 15 and 50 and glycine 16. Charged residues that form salt bridges in the HLA-A2 molecule, histidine 3 to aspartic acid 29, and arginine 111 to aspartic acid 129, are present in the mouse transplantation antigens and the TL^C molecule, while only the first is present in the $T9^C$ and $T18^C$ translations. Contact residues in the HLA-A2 molecule such as glutamine 115, which contacts β_2-microglobulin, as well as tryptophan 60, leucines 78, 160, 168, and 178, and alanine 153, which are α-helix residues that contact β-strand amino acids, are in general conserved either identically or with conservative substitutions. However, they are conserved to a greater degree in the transplantation antigens, possibly because they are more critical for antigen-presenting structures. Finally, two cysteines that form a disulfide bond in the α2 domains of all class I molecules, cysteines 101 and 164, are conserved in all of the aligned protein sequences.

Deletions and insertions present in the $T9^C$ and $T18^C$ translations suggest that if these genes or homologs in other mammals are expressed as proteins, their tertiary structure is probably at least somewhat different from those of the transplantation antigens. Whether deletions or insertions would radically alter the structure of these putative molecules would depend in part on where they are in relation to secondary structural characteristics of the transplantation antigens. Therefore, the placement of these deletions and insertions in the translations of

the $T9^C$ and $T18^C$ genes aligned to the HLA-A2 protein sequence was examined (Figure 3).

The $T9^C$ translation has three deletions relative to the HLA-A2 amino acid sequence, one in the α1 domain, and two in the α2 domain. All of these deletions reside within or near portions of the $T9^C$ translation that correspond to secondary structures in the HLA-A2 molecule. The first deletion corresponds to residues 46-48 within the fourth β strand of the α1 domain of the HLA-A2 molecule, while the second deletion corresponds to residues 136 and 137 of the fourth β strand of the α2 domain. The third deletion of eight amino acids corresponds to residues 151-158, which are in the α helix of the α2 domain in the HLA-A2 molecule. Although the helical portion of the α2 domain is 11 amino acids longer than the helical portion in the α1 domain in the HLA-A2 molecule, it is not clear whether a class I molecule with such an extensive deletion in the α2 domain helix could assume a transplantation antigen-like structure. Therefore, even if the $T9^C$ gene had an open reading frame in exon 3, its protein product would be missing at least 1 and probably 2 β strands critical to the α1 and α2 domain structure.

The amino acid sequence of the putative $T18^C$ molecule has four deletions and two insertions in the α1 and α2 domains relative to the HLA-A2 molecule. However, unlike those of the $T9^C$ translated amino acid sequence, these insertions and deletions generally appear to reside outside stretches that form critical secondary structures in the HLA-A2 molecule. The sole exception is a single amino acid deletion corresponding to residue 53 in the short helix of the α1 domain. The other deletions correspond to residues 39 and 40, 108, and 111, which reside between the β strands of the α1 and α2 domains of the HLA-A2 molecule. The two insertions are an 11 amino acid insertion between two residues homologous to 87 and 88 in the HLA-A2 molecule, and a single amino acid

insertion between residues that correspond to amino acids 150 and 151. These insertions are in loops between α helices and β strands in the HLA-A2 molecule, and thus it is possible that the insertions increase the size of these loops in the putative T18 molecule while not radically altering its structure from that of a transplantation antigen. Thus, unlike the *T9*^{c} gene, the *T18*^{c} gene appears to be capable of encoding a molecule that, in spite of deletions, insertions, and extensive amino acid sequence divergence, could possibly assume a structure similar to a transplantation antigen.

The structure assumed by class I α3 domains is believed to be an immunoglobulin homology-unit domain. The amino acid sequence and the size of the α3 domain are generally conserved between the classical class I molecules and all of the putative molecules encoded by the cDNA clones (Figure 3). In addition, three amino acids believed to be critical for the tertiary structure of immunoglobulin homology unit domains (35) are conserved in every predicted α3 domain. These residues are cysteines 203 and 259, which together form a disulfide bond as well as tryptophan 217, which is believed to form a critical intradomain contact. The size and general amino acid sequence conservation, as well as the conservation of these three residues, suggests that the putative non-classical class I molecules would also have immunoglobulin homology unit α3 domains.

Comparison of the predicted translations of exons 5, 6, 7 and 8 of the cDNA clones with the cytoplasmic domains of the D^{d} molecule reveals that those predicted to be encoded by *D* region genes are similar to the D^{d} molecule, while those of the *Tla* region genes share little similarity. The divergence of the cytoplasmic domains of the predicted *Tla* region molecules suggests that the cytoplasmic domains of these moleules would have functions different from those of the classical class I molecules. In the case of the putative T18^{c} molecule, it is

unclear how large the cytoplasmic domain is. The *T18*^{c} cDNA does not contain the complete 3' end, and there is a donor-splice sequence 5' of the end of the putative coding sequence in the cDNA, which could conceivably splice a 3' exon to the sixth exon of the *T18*^{c} gene. Since *T18*^{c} transcripts appear to be approximately 5000 nucleotides in length (Brorson *et al.* in preparation), theoretically they could encode molecules with a cytoplasmic domain much larger than those of the transplantation antigens. It has been suggested that conserved sequences within the cytoplasmic domains of classical class I molecules are required for their phosphorylation and endocytosis (36). Putative phosphorylation sites (S-$_{D}^{N}$-X-S-L) present in the cytoplasmic domains of classical class I molecules (37) are not present in the putative *Tla* region molecules, and can be identified only in the putative D3^{d} cytoplasmic domain. The absence of putative phosphorylation sites, which are conserved in classical class I molecules, suggests that the putative *Tla* region molecules may have different trafficking patterns within the cell than the transplantation antigens, or that the regulation of their appearance to and from the cell surface is different.

Evolutionary Relationships. The divergence of the putative *Tla* and *Hmt* region non-classical class I molecules suggests that some members of this family of molecules may be recognition structures that perform functions other than antigen presentation. Although it is unclear what functions they serve, their divergence suggests that they have been preserved in the genome for an extensive period of evolutionary history. Since many have maintained open reading frames over this extensive period, it is likely that selective pressure is operating to maintain them. If they have origins that date to early vertebrate evolution, it is possible that they may have a functional importance common to several types of vertebrates, such as in development. To study the evolution of the class I gene

family, rooted phylogenetic trees were constructed using a distance matrix method from synonymous substitution frequencies between the sequences in this report (Table Va and b), and other previously published class I sequences (Figures 4a and b).

The BALB/c genes included in this analysis were the *T9*^{c}, *T18*^{c}, and *D2*^{d} genes as well as two transplantation antigen genes, *H-2K*^{d} (34) and *D*^{d} (27), the *T13*^{c} gene that encodes the TL^C antigen (12), and the full-length *Thy19.4* gene (16). In addition, transplantation antigen genes from other mammals, rat *Ratmhc1* (38), human *HLA-A2* (33), miniature swine *PD1* (39), and rabbit *pR9* (40), were also included as well as the chicken *B-F* gene (41). Since the evolutionary history of these vertebrates has been established by the fossil record (42-45), their class I gene sequences can provide divergence time points with which the accuracy of this analysis can be tested.

In the comparisons, exons 2 and 3 were analyzed together, while exon 4 was analyzed separately because gene conversion events are suggested to occur frequently between fourth exons, making them appear to have diverged more recently than second and third exons (12,46). A synonymous substitution distance matrix method was chosen for this analysis instead of a parsimony method because there appears to be stronger selective pressure against coding region substitutions in some genes than in others (Table VI), and hence parsimony methods that analyze total nucleotide substitutions would generate a tree with artificially short branches for more stringently selected class I genes (data not shown). Introns were not compared because in many cases they are unalignable, or their sequences are unavailable. The UPGMA distance matrix method is used when comparing sequences with relatively high substitution frequencies (d) (48) as in the case of the second and third exon comparisons ($d > 1.00$ in some cases). Similar results

were obtained when the matrices were re-analyzed using the neighbor-joining method (67), with the *T9*^{c} and *T13*^{c} genes in addition to the *T18*^{c} and *Thy19.4* genes diverging from the transplantation antigen genes before bird/mammal divergence (data not shown).

This analysis reveals that the second and third exons of all four of the *Tla* and *Hmt* region genes, *T9*^{c}, *T13*^{c}, *T18*^{c}, and *Thy19.4* duplicated and diverged from the transplantation antigen genes prior to the mammalian radiation in the Cretaceous period 65-85 million years ago (42,45; Fig. 4a). This analysis correctly places the divergence of the modern Eutherian orders at 65-85 million years ago (42,45), and hence approximate divergence dates can be interpolated for genes that duplicated and diverged in the Cenozoic era. The oldest lineages appear to be those of the *Thy19.4* and *T18*^{c} genes, which appear to have duplicated and diverged from the transplantation antigens prior to bird/mammal divergence 300 million years ago in the Carboniferous period (42). Hence, by this analysis, it appears that these two genes date to at least the late-Paleozoic era. The *T9*^{c} and *T13*^{c} genes appear to have diverged from the transplantation antigen genes after bird/mammal divergence, and hence probably date to either the late Paleozoic or the Mesozoic eras. On the other hand, the classical transplantation antigen *H-2K*^{d} and *D*^{d} genes appear to have diverged from each other about 25 million years ago, approximately the time of mouse/rat divergence, estimated using molecular techniques (44,50). This analysis is consistent with a model of class I evolution which has proposed that *H-2/Tla* divergence was a primordial event that preceded the mammalian radiation, and that the divergence of *Qa* and *H-2* region genes occurred subsequently (61).

The evolution of the fourth exons of these genes appears to be markedly different (Figure 4b). Based on synonymous mutation frequencies, all of the

fourth exons appear to have diverged from a common ancestor after the divergence of the modern placental mammals. For each gene, the divergence of the fourth exon from the transplantation antigen lineage appears to had occurred subsequent to the divergence of the second and third exon. This is consistent with the hypothesis that gene conversion events had occurred between the fourth exons of these genes, creating a family of hybrid genes with fourth exons that are similar and second and third exons that are divergent (12). It is possible that gene conversions are favored to occur in the fourth exons because high selective pressure exists to maintain the structure of the α3 domains of class I molecules. On the other hand, it is conceivable that greater divergence is favored for the second and third exons because the domains that they encode do not have identical functions between all class I molecules. Divergence in the α1, α2 and cytoplasmic domain coupled with conservation of a class I-like structure in the α3 domain suggests that the putative molecules encoded by the *Tla* and *Hmt* region genes are β_2-microglobulin-associating cell-surface recognition structures, although probably not viral antigen-restriction elements.

This analysis also reveals that the *Tla* region genes as well as the classical class I genes have low ratios of synonymous to non-synonymous replacement mutations (Table VI). In most functional genes, the ratio of synonymous substitutions to non-synonymous substitutions between homologous genes ranges from approximately 2.5 to 360. However, in the class I genes, this ratio is generally less than 2.5, and in some cases, unity. It is possible that some of these genes are pseudogenes, and hence have no coding region selective pressure. However, the K^d and D^d genes encode molecules that are critical for the survival of mice, and the ratio of synonymous and non-synonymous mutations between them is unity. A low ratio could occur in homologous coding sequences that are

maintained by frequent gene conversion or duplication events, which is apparently the case in the fourth exons. In addition, it is possible that polymorphism enhances the function of some class I molecules, and hence little selective pressure would exist against replacement mutations (60). For example, a large populational diversity of *H-2* alleles can present a wide variety of antigens. The exception to the lack of selective pressure to maintain a particular coding sequence appears to be the *Thy19.4* gene, which has a synonymous-to-non-synonymous substitution ratio of 4.5 when compared to the $H\text{-}2D^d$ gene, possibly reflecting a moderate selective pressure to maintain a coding sequence giving rise to motifs that assume a structure similar to a transplantation antigen (16).

SUMMARY

This report reveals that 15 different class I genes are transcribed in the thymus, including several *Tla* region genes that have not been previously characterized. The analysis presented herein shows that the only extensive amino acid sequence similarity (and hence potential tertiary-structure similarity) between the putative *Tla* and *Hmt* region class I molecules and conventional transplantation antigens occurs exclusively in the β_2-microglobulin-binding $\alpha 3$ domain. The divergence of the predicted $\alpha 1$, $\alpha 2$, transmembrane, and cytoplasmic domains encoded by the *Tla* and *Hmt* region genes suggests that these domains perform functions other than those of the transplantation antigens, while the conservation of the $\alpha 3$ domain and a hydrophobic transmembrane segment is consistent with the hypothesis that these putative molecules are β_2-microglobulin-binding cell-surface molecules. Whatever function that the *Tla* and *Hmt* region genes have seems to be evolutionarily very old, and thus possibly fundamental, since some of these genes appear to have diverged from the classical class I genes early in vertebrate evolution. The thymus is the location of T-cell differentiation and contains a wide variety of hematopoietic cells as well as T cells in various stages of differentiation. Thus, the possible presence of class I cell surface structures in the thymus suggests the possibility that they are involved in cell-cell recognition events between different cell types present.

The authors wish to thank Drs. I. Stroynowski, M. Zuniga, and J. Kobori for critically reviewing this manuscript, T. Hunkapiller and Dr. B. Koop for critical insights on evolution and DNA sequence analysis, and Mrs. C. Blagg for expert secretarial assistance.

References

1. Silver, J. and L. Hood. 1974. Detergent solubilized H-2 alloantigen is associated with a small molecular weight polypeptide. *Nature* **249:** 764.
2. Klein, J. 1986. *Natural History of the Major Histocomptibility Complex.* John Wiley and Sons, New York, New York.
3. Zinkernagel, R. M. and P. C. Doherty. 1980. MHC-restricted cytotoxic T cells: Studies on the role of polymorphic major transplantation antigens determining T-cell restriction specificity, function and responsiveness. *Adv. Immunol.* **27:**51.
4. Old, L. J., E. A. Boyse, and E. Stockert. 1963. Antigenic properties of experimental leukemias. I. Serological studies *in vitro* with spontaneous and radiation-induced leukemias. *J. Natl. Cancer Inst.* **31:** 977.
5. Stanton, T. H., and E. A. Boyse. 1976. A new serologically defined locus, *Qa-1*, in the *Tla* region of the mouse. *Immunogenetics* **3:** 525.
6. Flaherty, L. 1976. The *Tla* region of the mouse: Identification of a new serologically defined locus, *Qa-2. Immunogenetics* **3:** 533.
7. Robinson, P. J. 1985. Qb-1, a new class I polypeptide encoded by the *Qa* region of the mouse *H-2* complex. *Immunogenetics* **22:**285.
8. 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.
9. Steinmetz, M., A. Winoto, K. Minard, and L. Hood. 1982. Clusters of genes encoding mouse transplantation antigens. *Cell* **28:**489.

10. 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 and evolution of the class I gene family in the major histocompatibility complex of the C57BL/10 mouse. *Nature* **310**:650.

11. 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.

12. Fisher, D. A., S. W. Hunt, and L. Hood. 1985. 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.

13. Stephan, D., H. Sun, K. Fischer Lindahl, E. Meyer, G. Hammerling, L. Hood, and M. Steinmetz. 1986. Organization and evolution of *D* region class I genes in the mouse major histocompatibility complex. *J. Exp. Med.* **163**:1227.

14. Richards, S., M. Bucan, K. Brorson, M. Kiefer, S. Hunt, H. Lehrrach, and K. Fischer Lindahl. 1989. Genetic and molecular mapping of the *Hmt* region of the mouse. *EMBO J.*, in press.

15. Stroynowski, I., M. Soloski, M. G. Low, and L. Hood. 1987. A single gene encodes soluble and membrane-bound forms of the major histocompatibility Qa-2 antigen: Anchoring of the product by a phospholipid tail. *Cell* **50**:759.

16. Brorson, K., S. Richards, S. W. Hunt, H. Cheroutre, K. Fischer Lindahl, and L. Hood. 1989. Analysis of a new class I gene mapping to the *Hmt* region of the mouse. *Immunogenetics*, in press.

17. 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.

18. Blobel, G. and B. Dobberstein. 1975. Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *J. Cell Biol.* **67**: 835.

19. Bjorkman, P. J., M. A. Saper, B. Samraoui, W. S. Bennett, J. L. Strominger, and D. C. Wiley. 1987. Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* **329**: 506.

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

21. Metcalf, D. 1966. The structure of the thymus. In *The Thymus: Recent Results in Cancer Research*, D. Metcalf, ed., Vol. 15, Springer-Verlag, Berlin, pp. 1-17.

22. Metheieson, B. J., and B. J. Fowlkes. 1984. Cell surface expression on thymocytes: Development and phenotypic differentiation of intrathymic subsets. *Immunological Revs.* **82**: 141.

23. Beller, D. I., and E. R. Unanue. 1978. Thymic macrophages modulate one stage of T-cell differentiation *in vitro*. *J. Immunol.* **121**: 1861.

24. Brorson, K. A., S. W. Hunt, T. Hunkapiller, Y. H. Sun, H. Cheroutre, D. Nickerson, and L. Hood. 1989. Comparison of exon 5 sequences from 35 class I genes of the BALB/c mouse. *J. Exp. Med.*, in press.

25. Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18:**5295.

26. Hyunh, T., R. Young, and R. Davis. Constructing and screening cDNA libraries in λgt10 and λgt11. In D. Glover (ed.) *DNA Cloning: A Practical Approach*, IRL Press, Oxford, 1984, pp. 49-78.

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

28. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. ***Molecular Cloning:*** *A* ***Laboratory Manual.*** Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

29. Feinberg, A. P., and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132:**6.

30. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74:**5463.

31. Robinson, P. J., D. Bevac, A. L. Mellor, and E. H. Weiss. 1988. Sequence of the mouse *Q4* class I gene and characterization of the gene product. *Immunogenetics* **27:**79.

32. Pampeno, C. L., and D. Meruelo. 1986. Isolation of a retrovirus-like sequence from the *Tla* locus of the C57BL/10 murine major histocompatibility complex. *J. Virol.* **58:**296.

33. Koller, B., and H. Orr. 1985. Cloning and complete sequence of an *HLA-A2* gene: analysis of two *HLA-A* alleles at the nucleotide level. *J. Immunol.* *134*: 2727.

34. Kvist, S., L. Roberts, and B. Dobberstein. 1983. Mouse histocompatibility genes: Structure and organziation of a K^d gene. *EMBO J.* **2**: 245.

35. Lesk, A., and C. Clothia. 1982. Evolution of proteins formed by β-sheets. II. The core of the immunoglobulin domains. *J. Mol. Biol. 160*: 325.

36. Capps, G., M. Van Kampen, C. Ward, and M. Zuniga. 1989. Endocytosis of the class I major histocompatibility antigen via a phorbol myristate acetate-induced pathway is a cell-specific phenomenon and requires the cytoplasmic domain. *J. Cell Biol. 108*: 1317.

37. McCluskey, J., L. Boyd, W. Maloy, J. Coligan, and D. Margulies. 1986. Alternative processing of *H-2D*^{d} pre-mRNAs result in membrane expression of differentially phosphorylated protein products. *EMBO J.* **5**: 2477.

38. Kastern, W. 1985. Characterization of two class I major histocompatibility rat cDNA clones, one of which contains a premature termination codon. *Gene 34*: 227.

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

40. Tykocinski, M., P. Marche, E. Max, and T. Kindt. 1984. Rabbit class I MHC genes: cDNA clones define full-length transcripts of an expressed gene and a putative pseudogene. *J. Immunol. 133*: 2261.

41. Guillemot, F., A. Billault, O. Pourquie, G. Behar, A. Chausse, R. Zoorob, G. Kreibich, and C. Auffray. 1988. A molecular map of the chicken major histocompatibility complex: the class II β genes are closely linked to the class I genes and the nucleolar organizer. *EMBO J.* **7**: 2775.

42. Young, J. Z. 1962. *The Life of Vertebrates.* Oxford University Press, New York.

43. Lillegraven, J., Z. Kielan-Laworowska, and W. Clemens. 1979. *Mesozoic Mammals, the First Two-Thirds of Mammalian History.* University of California Press, Berkeley.

44. Jaeger, J., H. Tong, and C. Denys. 1986. The age of *Mus rattus* divergence: paleontological data compared with the molecular clock. *C.R. Acad. Sc. Paris 302*: 917.

45. Gidley, J. 1912. The Lagomorphs an Independent Order. *Science 36*: 285.

46. Hayashida, H., and T. Miyata. 1983. Unusual evolutionary conservation and frequent DNA segment exchange in class I genes of the major histocompatibility complex. *Proc. Natl. Acad. Sci. USA 80*: 2671.

47. Jukes, T., and C. Cantor. 1969. Evolution of protein molecules. In H. N. Munro, ed. *Mammalian Protein Metabolism*, New York, Academic Press, pp. 21-132.

48. Nei, M. 1987. *Molecular Evolutionary Genetics.* Columbia University Press, New York.

49. Li, W.-H., C.-I. Wu, and C.-C. Luo. 1985. A new method for estimating synonymous and non-synonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol. 2*: 150.

50. Britten, R. 1986. Rates of DNA sequence evolution differ between taxonomic groups. *Science 231*: 1393.

51. 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.

52. 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.

53. Transy, G., R. S. Nash, B. David-Watine, M. Cochet, S. W. Hunt, III, L. Hood, and P. Kourilsky. 1987. A low polymorphic mouse *H-2* class I gene from the *Tla* complex is expressed in a broad variety of cell types. *J. Exp. Med.* **166**:341.

54. Hedley, M., S. Hunt, K. Brorson, J. Andris, L. Hood, J. Forman, and P. Tucker. 1989. Analysis of $D2^d$ a *D* region class I gene. *Immunogenetics* **29**:359.

55. Fisher, D., M. Pecht, and L. Hood. 1989. DNA sequence of a class I pseudogene from the *Tla* region of the murine MHC: Recombination of a B2 Alu repetitive sequence. *J. Mol. Evol. 28*: 306.

56. Widmark, E., H. Ronne, U. Hammerling, R. Servenius, D. Larhammar, K. Gustafsson, J. Bohme, P. Peterson, and L. Rask. 1988. Family relationships of murine major histocompatibility complex class I genes. Sequence of the $T2A^a$ pseudogene, a member of gene family 3. *J. Biol. Chem. 263*: 7055.

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

58. Matsuura, A., R. Schloss, F.-W. Shen, J. Tung, S. Hunt, D. Fisher, L. Hood, and E. Boyse. 1989. Expression of the $Q8/9^d$ gene by inducer-T cells of the mouse, in preparation.

59. Kimura, M. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. *Proc. Natl. Acad. Sci. USA 78*: 454.

60. Arden, B. and J. Klein. 1982. Biochemical comparison of major histocompatibility complex molecules from different subspecies of *Mus musculus*: Evidence for trans-specific evolution of alleles. *Proc. Natl. Acad. Sci. USA 79*: 2342.

61. Klein, J. and F. Figueroa. 1986. The evolution of class I MHC genes. *Immunology Today* 7:41.

62. Goodenow, R. S., M. McMillan, M. Nicholson, 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.

63. Hood, L., M. Steinmetz, and B. Malissen. 1983. Genes of the major histocompatibility complex. *Annu. Rev. Immunol.* 1:259.

64. Mount, S. M. 1982. A catalogue of splice junction sequences. *Nucl. Acids Res.* 10:459.

65. Hubbard, S. and R. Ivatt. 1981. Synthesis and processing of asparagine-linked oligo saccharides. *Ann. Rev. Biochem.* 50: 555.

66. Stroynowski, I. 1990. Molecules related to class I major histocompatibility antigen. *Ann. Rev. Immunol.*, in press.

67. Saitou, N. and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4:406.

Table I

Current Status of the 35 Class I Genes Present in BALB/c Mice

Gene	Product	Expression	Reference
K^d	H-2K^d	Entire body	34
D^d	H-2D^d	Entire body	27
L^d	H-2L^d	Entire body	51
$D2^d$	?	?	54
Q4	Qb-1 soluble class I	Entire body	31
$Q7^d$	Qa-2 antigen	Lymphoid	17
Q10	Q10 soluble class I	Liver	52
$T13^c$	TL antigen	Thymus	12
37^c	?	Entire body	53
$T1^c$	ψ gene	-	55
T2	ψ gene	-	56
$T4^c$	ψ gene or secreted	-	24
$T5^c$	ψ gene	-	57
$T10^c$	ψ gene	-	D. Nickerson, per. comm.
$T11^c$	ψ gene or secreted	-	24
$T12^c$	ψ gene or secreted	-	24
Thy19.4	?	Entire body	16

The *Q4*, *Q10*, and *T2* genes were sequenced in other haplotypes of mice. The site of expression of the class I pseudogenes has not been determined.

Table II
Class I Gene Transcripts Identified in the cDNA Library from Balb/c Thymus

Gene	MHC Location	Number of Clones
K^d	*K*	21
D^d	*D*	11
$D2^d$	*D*	2
$D3^d$	*D*	2
$D4^d$	*D*	1
L^d	*D*	7
$Q4^d$	*Qa*	3
$Q7^d$	*Qa*	2
$Q8/9^d$	*Qa*	2
$T9^c$	*Tla*	1
$T13^c$	*Tla*	4
$T17^c$	*Tla*	5
$T18^c$	*Tla*	1
37^c	*Tla*	5
Thy19.4	*Hmt*	2
unidentified		3
	Total	72

Those that are unidentified do not contain the fifth exon used to identify clones in this report.

Table III
Splicing Characteristics of cDNA Clones

Gene	cDNA Clone Size	Exons Contained	Introns Contained	Splice?[1]
$D2^d$	1500 bp	/2,3,4,5,6,7,8	6	yes
			7	no
$D3^d$	1406 bp	/4,5,6,7,8	3,5,6,7	yes
$D4^d$	1095 bp	/4,5,6,7,8	3,7	yes
$Q7^d$	1509 bp	/4,5,6,8	3,4,5	yes
$Q8/9^d$	1970 bp	/4,5,6,7,8	3,4,5,6,7	yes
$T9^c$	1931 bp	1,2,3,4,5,6	2	yes
$T17^c$	913 bp	/4,5,6		
$T18^c$	2125 bp	1,2,3,4,5,6/		
Thy19.4	~ 1500 bp[5]	/4,5/	3,4	yes

1. In some cases, whether these transcripts can splice out the introns is surmised from the presence of consensus splice sites (64). In other instances, splicing patterns were determined by comparison to cDNA clones from other studies. A $D2^d$ cDNA clone derived from a T-inducer cell cDNA library does not contain intron 6 (data not shown), and hence it is possible that this intron is normally removed from $D2^d$ transcripts. On the other hand, there is no donor-splice sequence at the predicted location near exon 7 of the $D2^d$ transcript, and thus it is unlikely that any splicing occurs 3' of exon 7. The $Q8/9^d$ exon 7 acceptor-splice signal contains several purines in the polypyrimidine tract 5' of the splice junction and thus appears non-functional. Since the $Q7^d$ transcript has spliced exon 6 to exon 8 instead of exon 7, it is likely that $Q8/9^d$ transcripts splice these two exons to each other as well. It is likely that introns 4 and 5 are spliced

from $Q7^d$ and $Q8/9^d$ transcripts since transcripts from the $Q7^b$ gene have been shown to remove these introns (58). Finally, the splicing pattern of the *Thy19.4* gene was determined in another study (16).

2. The first 156 bp of the $D2^d$ cDNA are similar to the complement of exons 1 and 2 of the $H\text{-}2D^d$ gene. Since the 5' end of the $D2^d$ gene has a typical class I exon/intron structure (54), it is likely that the sequence inversion is a cloning artifact.

3. The number of exons that contribute to the 3' end of transcripts from these genes is unclear since the 3' sequences of these genes share little nucleotide sequence similarity with characterized genes.

4. Nucleotides 1-57 of the $T18^c$ cDNA are exon 1 sequences. It is unclear whether this sequence corresponds to the entire first exon, or if this clone was truncated at the 5' end.

5. Only 819 bp of the *Thy19.4* cDNA sequence is included in this study. The excluded sequence is from intron 3 and the 3' untranslated portion.

TABLE IVa

Percent Nucleotide Sequence Similarity of Class I Genes with H-2D^d

Gene	*H-2K^d*	*D2^d*	*D3^d*	*D4^d*	*Q7^d*	*Q8/9^d*	*T9^c*	*T17^c*	*T13^c*	*T18^c*	*Thy19.4*
exon 1	86%								64%		53%
exon 2	89%	87%					67%		72%	53%	67%
exon 3	91%	87%					64%		74%	59%	68%
exon 4	93%	95%	95%	97%	93%	93%	93%	94%	92%	82%	88%
exon 5	82%	82%	84%	85%	81%	81%	39%	39%	65%	32%	52%
exon 6	94%	91%	88%	88%	88%	88%					
exon 7	92%	82%	92%	92%		87%					

TABLE IVb

Percent Nucleotide Sequence Similarity of cDNA Clones from Gene Pairs

	Q8/9^d vs. *Q7^d*	*T9^c* vs. *T17^c*
exon 4	96%	94%
exon 5	99%	99%
exon 6	100%	

For reference, the *H-2D^d* and *H-2K^d* genes are compared to each other. The nucleotide sequences of the *H-2D^d* (27), *H-2K^d* (34), *T18^c* (12), and *Thy19.4* exons 2 and 3 (16) were determined previously. A gap of any size is counted as one mismatch.

Table Va
Exons 2 and 3 Distance Matrix

D^d	.11	.19	.30	.35	.35	.57	.61	.68	1.11	1.16
K^d		.18	.38	.38	.42	.66	.57	.71	.97	1.31
D^d			.47	.47	.47	.55	.76	.59	.71	.97
Hum				.23	.29	.61	.88	.87	1.11	.88
Pig					.29	.59	.82	.73	.88	.95
Rab						.61	.82	.71	.79	1.02
$T9^c$							.79	.94	1.11	.88
$T13^c$								.94	.88	.92
Chi									1.11	1.43
$T18^c$										1.31

	K^d	$D2^d$	Hum	Pig	Rab	$T9^c$	$T13^c$	Chi	$T18^c$	*Thy19.4*
x (mean)	.11	.18	.38	.36	.36	.60	.75	.77	.96	1.10
σ (std. dev.)	-	.01	.07	.09	.07	.03	.11	.12	.15	.21
my	24	39±2	82±15	78±20	78±15	-	-	300	-	-

The values are synonymous replacement frequencies corrected using a binomial formula d = -.75 x ln(1-4p/3) (47). To prepare this matrix for analysis using the unweighted pair-group with arithmetic mean method (UPGMA), the genes are ordered in the matrix by distance from the *H-2D*^{d} gene. The average distance from *H-2D*^{d} (x) and a standard deviation (σ) are shown on the bottom of the matrix, as is the corresponding divergence time from the *H-2D*^{d} gene in millions of years (my). The human *HLA-A2* gene is abbreviated as Hum, Pig *PD1* as pig, rabbit *PR9* as rab, chicken *B-F* as Chi, and rat *ratmhc1* as rat.

Table Vb
Exon 4 Distance Matrix

D^d	.06	.10	.11	.08	.18	.27	.33	.30	.30	.99
K^d		.10	.13	.12	.12	.35	.33	.38	.33	1.02
$T9^c$			.12	.12	.18	.35	.40	.38	.38	1.11
Thy19.4				.14	.18	.37	.33	.37	.35	1.25
Rat					.15	.30	.40	.37	.40	.99
$T13^c$						.35	.43	.43	.40	1.15
$T18^c$							.47	.47	.61	1.15
Hum								.33	.22	1.02
Pig									.40	.79
Rab										.71

	K^d	$T9^c$	*Thy19.4*	Rat	$T13^c$	$T18^c$	Hum	Pig	Rab	Chi
x (mean)	.06	.10	.12	.12	.16	.33	.38	.38	.37	1.02
σ (std. dev.)	-	0	.01	.02	.02	.03	.05	.05	.10	.16
my	13	22	26±2	26±4	35±4	72±6	82±11	82±11	80±22	300

Table VI

Comparison of Synonymous to Non-Synonymous Frequency Ratios Among Various Mammalian Genes and Class I Genes

Genes Compared	Silent/Replacement Rates
Mammalian H4 Histones	357
Mammalian α-actins	262
Mammalian fibrinogens	10.6
Mammalian β_2-microglobulins	9.72
Mammalian α-fetoprotein	4.05
Mammalian β-globins	3.40
Mammalian α1 interferons	2.50
Exons 2 & 3 *Thy19.4* vs. D^d	4.5
Exons 2 & 3 $T13^c$ vs. D^d	2.3
Exons 2 & 3 $T18^c$ vs. D^d	2.1
Exons 2 & 3 $T9^c$ vs. D^d	1.4
Exons 2 & 3 *HLA-A2* vs. D^d	1.4
Exons 2 & 3 *B-F* vs. D^d	1.4
Exons 2 & 3 K^d vs. D^d	0.92
Exon 4 *Thy19.4* vs. D^d	1.0
Exon 4 $T13^c$ vs. D^d	3.0
Exon 4 $T18^c$ vs. D^d	1.9
Exon 4 $T9^c$ vs. D^d	1.7
Exon 4 *HLA-A2* vs. D^d	2.2
Exon 4 *B-F* vs. D^d	1.4
Exon 4 K^d vs. D^d	0.86
Exon 4 *Ratmhc1* vs. D^d	2.0

Silent and replacement frequencies p are corrected by a binomial formula $d = -3/4 \ln (1-4p/3)$ (47) and divided by each other. The values derived from non-class I

genes were adapted from Li *et al.* (1985).

Fig. 1. ***Map of class I genes in the BALB/c MHC.*** Class I genes map to the *K*, *D*, *Qa*, *Tla* and *Hmt* regions. The *I* and *S* regions contain class II and complement genes, respectively. The order of the *Tla* region gene clusters is unknown, as is the distance between the *K*, *D*, *Tla*, and *Hmt* regions. The upper line represents the genetic map and the gene clusters are indicated below.

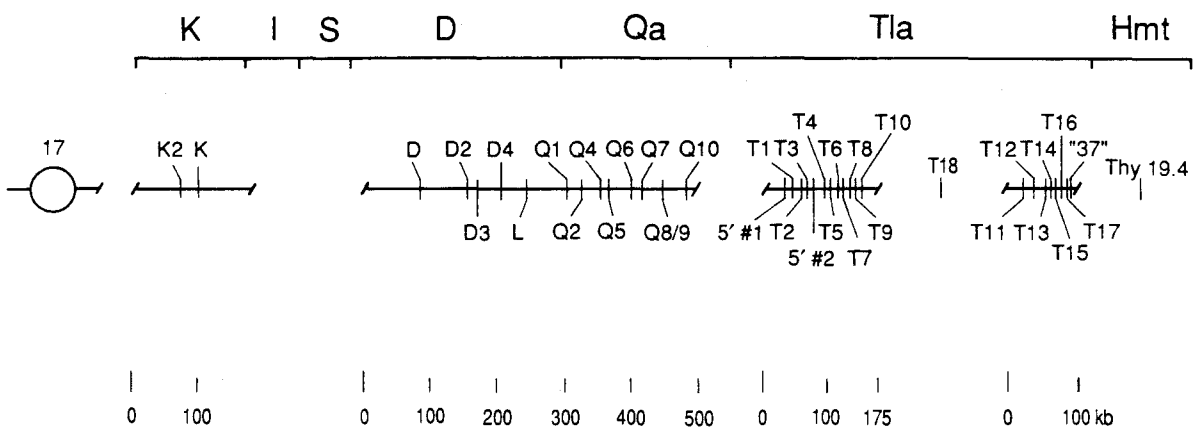
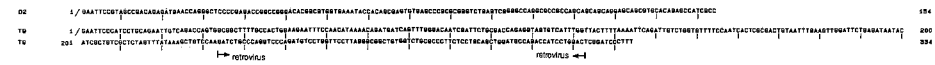

K
I
S
D
Qa
Tla
Hmt
17
K2 K
D D2 D4 Q1 Q4 Q6 Q7 Q10
D3 L Q2 Q5 Q8/9
T4 T10
T1 T3 T6 T8
5′ #1 T2 T5 T9
5′ #2 T7
T18
T16
T12 T14 "37"
Thy 19.4
T11 T13 T17
T15
0 100
0 100 200 300 400 500
0 100 175
0 100 kb

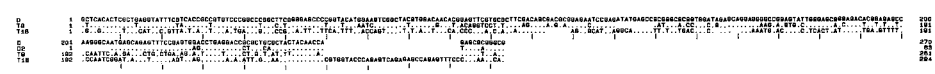
Fig. 2. *DNA sequences of the cDNA clones.* Where possible, the sequences are aligned to the *H-2D*^{d} gene sequence (27) or to each other. 5' sequences include a possible *T18*^{c} exon 1 sequence, and inverted *D2*^{d} exon 2 sequences. Retroviral sequences interrupting *T9*^{c} first exon are indicated below sequence. In the 3' sequences, exon 5, 6, 7, and 8 boundaries are indicated when possible. Potential splice signals in the 3' sequences of the *T9*^{c}, *T17*^{c}, and *T18*^{c} cDNA sequences are marked by arrows.

5′ Sequences



Exon 1

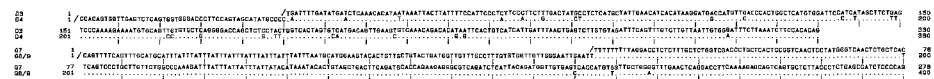
Exon 2



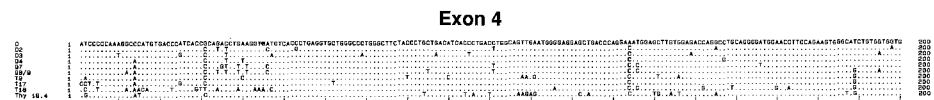
Intron 2

Exon 3

Intron 3



Exon 4



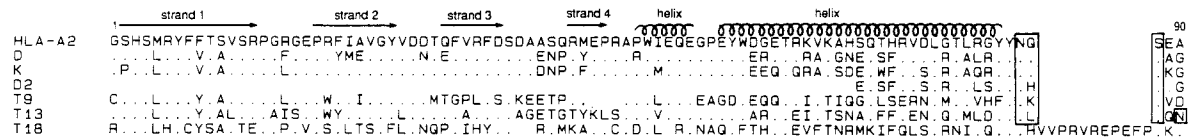
Intron 4

3′ Exons

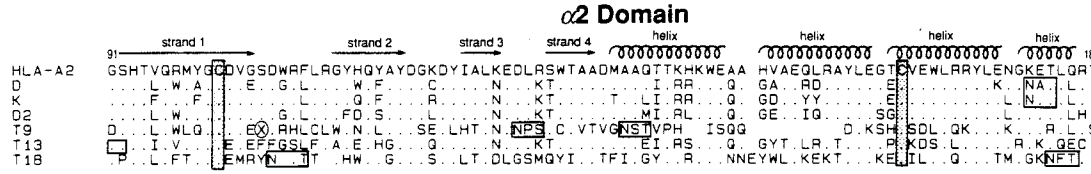
Fig. 3. *Aligned translations of cDNA clones.* In the α1, α2 and α3 domains, each sequence is aligned to that of the HLA-A2 molecule (33). The protein sequences of the K^d (34) and TL^c (12) molecules are also shown. Secondary structure characteristics of the HLA-A2 molecule are shown above its sequence (19). In the transmembrane and cytoplasmic domains, some of the sequences are aligned to the D^d sequence (27), while others are unalignable and shown individually. Possible donor-splice signals exist in the 3' portion of the $T9^c$, $T17^c$ and $T18^c$ cDNA sequences before the termination of translation, and hence it is unclear whether the translation of the cDNA beyond the signal is of exon or intron sequence. Potential glycosylation sites in the external domains and phosphorylation sites in the cytoplasmic domains are boxed (37,65). Invariant immunoglobulin homology-unit domain cysteines and tryptophans are shaded. The codon containing a single nucleotide deletion present in the third exon of the $T9^c$ gene is translated as an X. The seventh exon of the $D2^d$ gene has an eight nucleotide deletion, and hence the putative cytoplasmic domain that it encodes is nine amino acids shorter than that of the D^d molecule.

Leader Segments

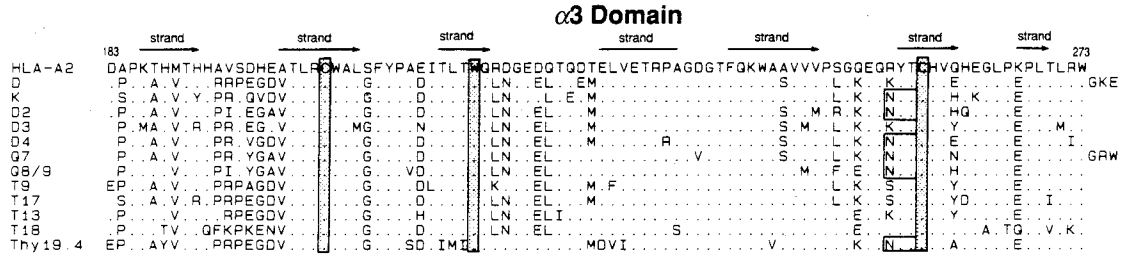
α1 Domain



α2 Domain



α3 Domain



Transmembrane and Cytoplastic Domains

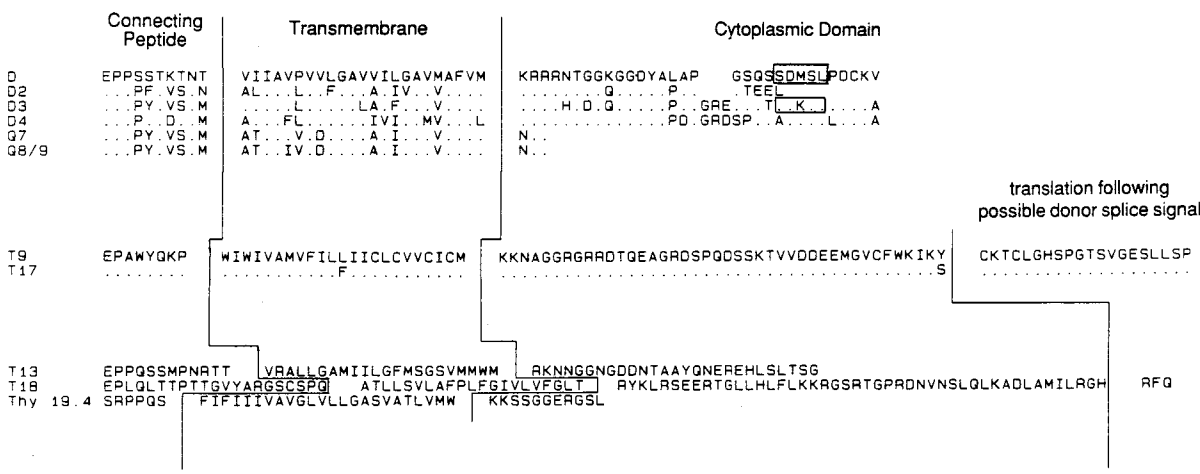
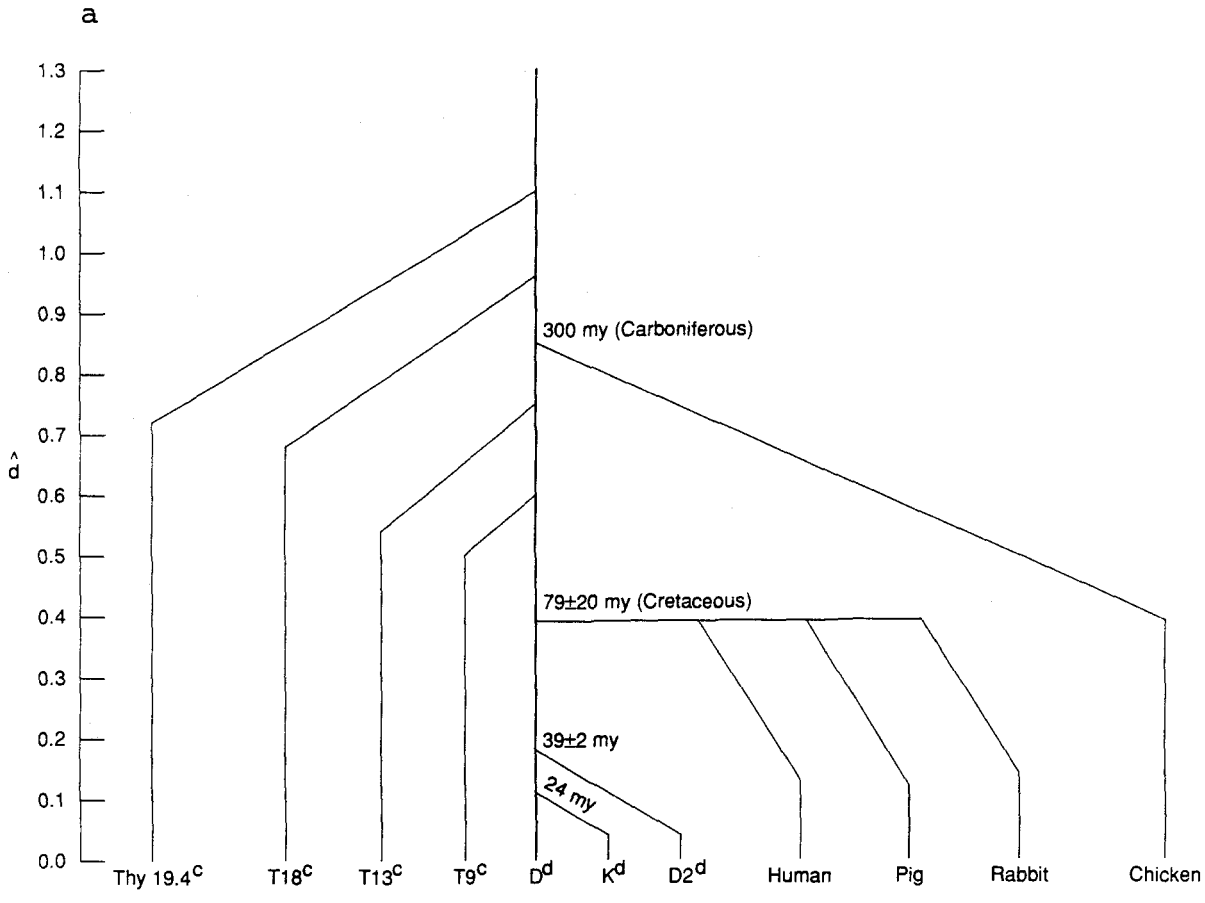
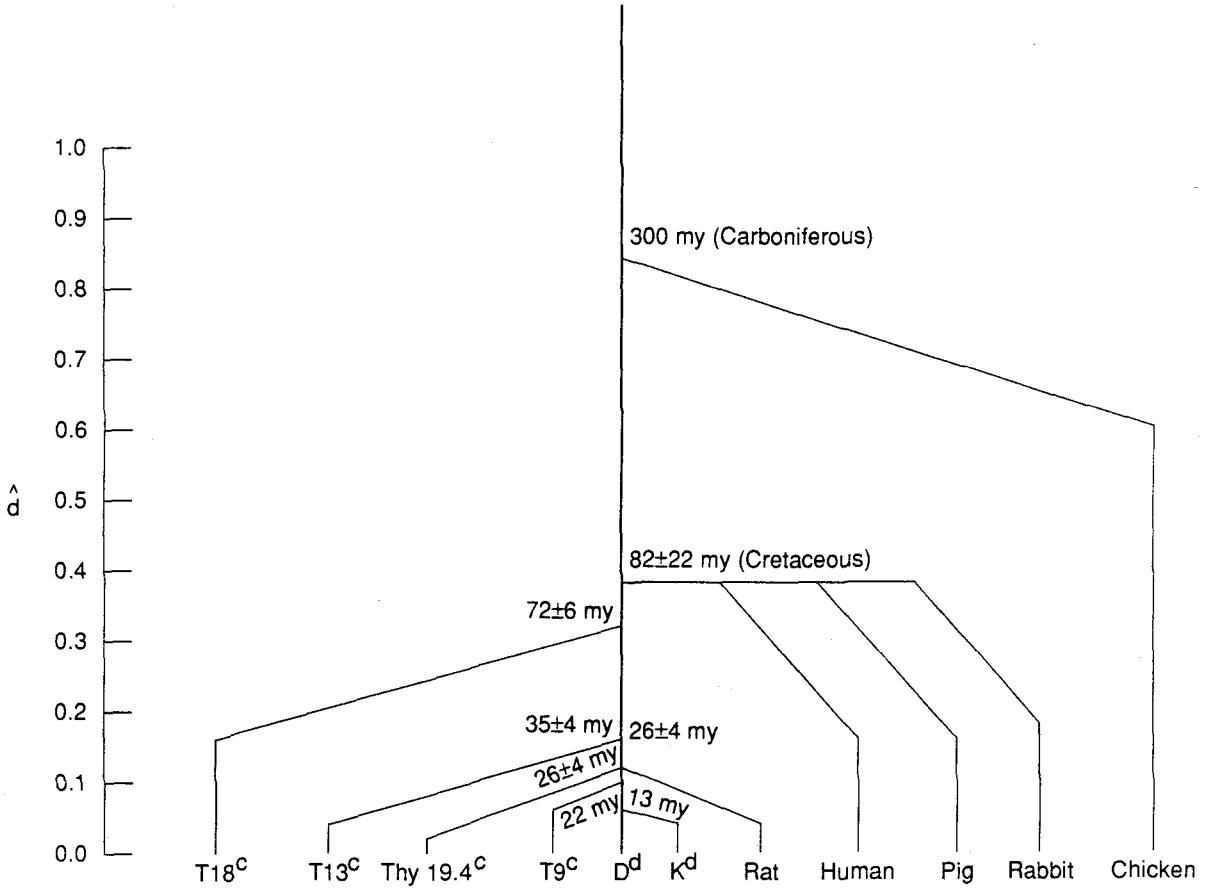


Fig. 4. (a) *Rooted phylogenetic tree constructed from exon 2 and 3 distance matrix.* This tree was constructed using the unweighted pair-group with the arithmetic mean (UPGMA) method (48). Since synonymous mutations are proposed to occur in mammals at a rate of approximately 4.6×10^{-9} substitutions per site per year, lengths in years of branches occurring after the mammalian radiation can be calculated by dividing average distance values (d) by 4.6×10^{-9} (49). The d value of the chicken *B-F* gene is an average of exons 2, 3 and 4.

(b) *Rooted phylogenetic tree constructed from exon 4 distance matrix.*


a
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
d̂
300 my (Carboniferous)
79±20 my (Cretaceous)
39±2 my
24 my
Thy 19.4c
T18c
T13c
T9c
Dd
Kd
D2d
Human
Pig
Rabbit
Chicken

b


1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
d̂
300 my (Carboniferous)
82±22 my (Cretaceous)
72±6 my
35±4 my
26±4 my
26±4 my
22 my
13 my
T18c
T13c
Thy 19.4c
T9c
Dd
Kd
Rat
Human
Pig
Rabbit
Chicken

Chapter 3

Analysis of a New Class I Gene Mapping to the *Hmt* Region of the Mouse

In press at *Immunogenetics*

Analysis of a New Class I Gene Mapping to the *Hmt* Region of the Mouse

Kurt A. Brorson[1], Sue Richards[2,5] Stephen W. Hunt III[1,3], Hilde Cheroutre[1,4], Kirsten Fischer Lindahl[2] and Leroy Hood[1]

1 Division of Biology, 147-75, California Institute of Technology, Pasadena, CA 91125, U.S.A.

2 Howard Hughes Medical Institute and Departments of Microbiology and Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX 75235-9050, U.S.A.

3 Present address: Department of Medicine, University of North Carolina, Chapel Hill, NC 27599, U.S.A.

4 Present address: Department of Microbiology and Immunology, University of California, Los Angeles, CA 90024, U.S.A.

5 Present address: Gene Screen Corp., Dallas, TX 75207, U.S.A.

Abstract. The major histocompatibility complex (MHC) of the BALB/c mouse contains three genes encoding classical class I molecules, as well as at least 32 non-classical class I genes. Although much is known about the genes encoding the classical class I molecules, the majority of the non-classical genes have not been characterized. This report describes a newly identified non-classical class I gene, *Thy19.4*, which contains an open reading frame and resembles, in several respects, the genes encoding classical class I molecules. The similarities include shared amino acid sequence motifs, which suggest that the putative Thy19.4 molecule may assume a tertiary structure similar to that of the classical class I molecules, as well as widespread transcription in a variety of tissues. However, unlike the classical class I genes, the *Thy19.4* gene maps approximately 1 centimorgan distal to the *Tla* region of the MHC, in the same region as the gene encoding the *Hmt* element of the maternally transmitted antigen.

Introduction

The class I molecules of the major histocompatibility complex (MHC) are cell-surface glycoproteins with Mr of approximately 45,000 that associate non-covalently with β_2-microglobulin, an Mr 12,000 polypeptide (Silver and Hood 1974). In the mouse, there are two types of class I molecules, the classical class I molecules, H-2K, D and L, encoded within the *H-2* region and the non-classical class I molecules encoded within the *Qa* and *Tla* regions. The classical class I molecules serve as viral antigen restriction elements for cytotoxic T lymphocytes (CTLs; Zinkernagel and Doherty 1980) and are present on virtually all nucleated cells (Klein 1975). The non-classical molecules are encoded by genes in the *Qa/Tla* region. They include the Qa-1 (Stanton and Boyse 1976), Qa-2 (Flaherty 1976), Qb-1 (Robinson 1985), and TL (Old *et al.* 1963) molecules, which are expressed in a tissue-specific manner and have no known function.

The molecular cloning of the genes encoding the murine class I (as well as II and III) molecules allowed the development of a detailed map of the MHC (Steinmetz *et al.* 1982; Weiss *et al.* 1984). In the BALB/c mouse there are many more class I genes, on the order of 35, than there are defined molecules (Fig. 1). The majority of the class I genes reside in the *Qa/Tla* region of the mouse MHC (Winoto *et al.* 1983), a region that is distal to the classical *H-2* loci on chromosome 17. The relatively large number of non-classical class I genes has prompted various attempts to determine which of these genes are expressed, and whether they are capable of encoding class I products. These studies have included sequence analyses of multiple class I cDNA clones (Hunt *et al.* in preparation; Lallane *et al.* 1985), and entire class I genes (Fisher *et al.* 1985; Transy *et al.* 1987). While these efforts have identified a few non-classical class I genes, both those with open reading frames and those without, the majority of the

32 non-classical BALB/c class I genes have not been characterized to any degree. In this report, we describe a newly identified non-classical class I gene, *Thy19.4*, which is the first well-characterized class I gene residing in a new region of the MHC, the *Hmt* region (Richards *et al.* 1989).

Materials and Methods

Isolation of Thy19.4 cDNA clone. A cDNA library was made from RNA isolated from the thymuses of four-week-old BALB/cJ mice using standard technologies (Huynh *et al.* 1984). It was screened with a probe derived from the fourth exon of the *H-2D*^{d} gene (Sher *et al.* 1985). Two of the 69 class I clones were *Thy19.4* transcripts. One of these clones, 19.4, was sequenced completely on both strands (Hunt *et al.* in preparation).

Isolation of Thy19.4 genomic clone. BALB/cJ liver DNA was isolated using standard protocols (Maniatis *et al.* 1982), digested with HindIII, and electrophoresed on an 0.8% agarose horizontal gel. The 7.2 kilobase (kb) fraction was isolated using an Elutrap device (Schleicher and Schuell, Kenne, NH). The insert was ligated into the HindIII site of phage λ762 DNA, and C600 ΔHfl *E. coli* bacteria were infected with the packaged ligation product. The resulting library was plated and screened under stringent conditions (68°C, 0.2xSSC) using a 3' probe derived from the *Thy19.4* cDNA clone (Fig. 2). Positive clones were plaque purified, and phage DNA was prepared as described previously (Maniatis *et al.* 1982).

Sequencing of the Thy19.4 clone. Four restriction enzyme fragments from the *Thy19.4* λ clone were subcloned into the M13 vectors mp18 and mp19. These

fragments included 1) the two 0.6 kb EcoRI/HindIII fragments containing the L1 repeats at either end of the 7.2 kb HindIII insert; 2) the 1.3 kb EcoRI fragment containing the 5' portion of the *Thy19.4* gene; and 3) the 2.4 kb EcoRI fragment containing the 3' portion of the *Thy19.4* gene (Fig. 2). With the exception of the 3'-end L1 repeat, these subclones were completely sequenced on both strands by the dideoxynucleotide chain termination technique (Sanger *et al.* 1980), using M13 and internal primers. The 3' EcoRI/HindIII fragment was confirmed as an L1 repeat by a single sequencing run from the 5' end.

DNA hybridizations. Wild mouse DNA was prepared by Dr. Rick Barth (Rochester, NY) from mice obtained from Dr. Verne Chapman (Buffalo, NY) (Fig. 3). Brno, Denmark, and Belgrade refer to the geographic area where the *Mus musculus musculus* mice were caught. Recombinant mouse DNA was prepared from mice bred in the colony of Dr. Kirsten Fischer Lindahl (Richards *et al.* 1989; Fig. 6b). The DNA was digested with PstI, or BamHI and electrophoresed in a 0.8% agarose horizontal gel. The DNA was transferred onto a Zeta probe membrane (Bio-Rad Co., Richmond, CA), and probed with the *Thy19.4* 3' probe. The filters were washed under high stringency (68°C, 0.2xSSC) and exposed for autoradiography for 16 h.

Polymerase chain reaction (PCR). Oligo-dT-primed single-stranded cDNA was synthesized using standard techniques (Hyunh *et al.* 1984) from 0.2 μg of total RNA from each of nine different tissues and four fetuses as well as total RNA from 7 cell lines obtained from ATCC (Rockville, MD): WEHI-3 (Metcalf *et al.* 1969); I-10 (Yasumura *et al.* 1966); BALB/c CL.7 (Patek *et al.* 1978); BNL CL.2 (Patek *et al.* 1978); J558 (Lundblad *et al.* 1972); J774 A.1 (Ralph and Nakoinz 1973); and

BALB/3T3 (Aaronson and Todaro 1968). Prior to cDNA synthesis, all RNA preparations were pretreated with RNAse-free DNAse (Worthington Biochemical Co., Freehold, NJ) to eliminate contaminating genomic DNA. PCR amplification was conducted in a total volume of 100 µl using TaqI polymerase (Cetus Inc., Emeryville, CA), and a thermal cycling machine for 25 cycles with an annealing temperature of 60°C (Saiki *et al.* 1988). Seven oligonucleotides were used to amplify four segments of *Thy19.4* sequences from the cDNA preparations, and genomic DNA. Exons 1, 2, and 3 were amplified using exon 1 (5' TCACTCTCTACCTGCTGCTTGGGCT 3', nt 799-823), and exon 3 (5' TCCAGGTGCTCAGATCTTCATTC 3', nt 1667-1645) oligonucleotide primers from genomic DNA, and all of the cDNA preparations. Exons 3, 4 and 5 were amplified using an exon 3 primer (5' GAATGAAGATCTGAGCACCTGGA 3', nt 1645-1667) and an exon 5 primer (5' CAGGCCAACAGCAACTATTATGAT 3', nt 2987-2964) from genomic DNA and thymus cDNA. Finally, genomic DNA and thymus cDNA was amplified using an exon 5 oligonucleotide primer (5' ATCATAATAGTTGCTGTTGGCCTG 3', nt 2964-2987) and two oligonucleotide primers (5' TTTATTAATATTCCCAGAAGACTTA 3', nt 3640-3616) and 5' TTTATTTCATTCTTACATCATTTCT 3', nt 3909-3885) derived from immediately 5' of the two poly(A) addition signals. After amplification, 10 µl of the amplified cDNA was electrophoresed on a 2% agarose gel, transferred to nitrocellulose, and hybridized with the *Thy19.4* 5' or 3' probes (Fig. 2) labeled by random hexanucleotide priming (Feinberg and Vogelstein 1983). The filters were washed at high stringency (68°C, 0.2X SSC), and exposed for autoradiography, using Cronex Lightning Plus intensifying screens (Dupont Co., Wilmington, DE).

Results

Isolation of Thy19.4 cDNA. A *Thy19.4* gene transcript was originally isolated from a thymus cDNA library, made from four-week-old BALB/cJ mice with a probe containing the conserved fourth exon of a class I gene (Hunt *et al.* in preparation). DNA sequence analysis of the clone indicated that its 3' sequences are similar to those of other class I genes, but did not match precisely any of the six previously published class I sequences (Steinmetz *et al.* 1981; Moore *et al.* 1982; Kvist *et al.* 1983; Fisher *et al.* 1985; Sher *et al.* 1985; Transy *et al.* 1987). The clone contained the fourth and fifth exons of a *Thy19.4* transcript as well as the third and fourth introns and sequences 3' of exon 5.

A gene-specific probe was generated by isolating the fifth exon and 3' sequences of the cDNA on a 550 basepair (bp) BglII/EcoRI fragment. Southern blot analysis using this probe demonstrates that the gene exists in a variety of rodents including rat and Peromyscus mouse (Fig. 3). Interestingly, only distantly related non-*musculus* species display size polymorphism from the 3.5 kb BamHI band that hybridizes to the 3' probe in *domesticus* strains.

Isolation of the Thy19.4 Genomic Clone. Hybridization experiments with the *Thy19.4* 3' probe demonstrated that the *Thy19.4* gene does not reside in any representative from a panel of class I MHC cosmids that span the *H-2*, *Qa*, and *Tla* regions (data not shown; Steinmetz *et al.* 1982). To isolate a *Thy19.4* genomic clone, a 7.2 kb HindIII size-selected insert library was made from BALB/cJ liver DNA because a fragment of that size hybridizes to the *Thy19.4* 3' probe (data not shown). One clone was isolated and mapped using BamHI, EcoRI, and BglII (Fig. 2). The clone contains the entire *Thy19.4* gene, as well as two L1 long interspersed repetitive sequences (LINES).

Sequence of the Thy19.4 Gene. The complete sequence was determined from the 5' HindIII site to the third EcoRI site (Figs. 2 and 4). An L1 repeat is included in the first 329 nucleotides of this sequence. It differs by just 1 basepair from the previously sequenced L1 repeat 3' of the mouse β-globin pseudogene (Hutchison *et al.* 1984).

Open reading frames corresponding to exons 1, 2, 3, 4, and 5 can be identified within this sequence by nucleotide-sequence similarity with other class I genes. The RNA splice sequences conform well to the consensus sequence defined by analysis of 139 exon-intron boundaries (Mount 1982). As in all eukaryotic genes, the GT/AG rule is followed at the immediate splice junctions (Breathnach and Chambon 1981). The nucleotide match of the four donor sequences to the splice-donor consensus sequence ($^{C}_{A}$AG/GT$^{A}_{G}$AGT) is greater than or equal to 5, the lowest value found for functional donor sequences. In the four acceptor-splice sequences, only that of exon 3 lacks a polypyrimidine tract normally found before the splice junction. However, most mouse class I genes also lack that particular polypyrimidine tract and still splice exon 2 to exon 3. In addition, PCR amplification experiments described later in this article demonstrate that *Thy19.4* gene transcripts splice all five exons to each other correctly.

Exactly 109 bp 5' of the initiating codon is a CAAT box, and a transcription initiation TATA box is 48 bp 5' of the initiating ATG codon. The 44 nucleotides between the TATA box and the initiating ATG codon places the start of transcription at least 10 nucleotides 5' of the putative initiation of translation, a distance that is typical of eukaryotic genes (Breathnach and Chambon 1981; Kozak 1984). In addition, most class I TATA boxes are found in roughly the same position (Kvist *et al.* 1983; Israel *et al.* 1986). The *Thy19.4* CAAT box, on the other hand,

is 20 bp 5' of the position where they are normally found in class I promoter regions. An interferon regulatory sequence, AGTTTCCCTTCT, is found 168 bp 5' of the initiation codon. This position is similar to those of interferon regulatory sequences in other class I genes (Israel *et al.* 1986; Korber *et al.* 1988). Intron 3, at 720 nucleotides, is short compared to the 1.7 kb third introns of the $H\text{-}2D^d$ and $H\text{-}2K^d$ genes (Sher *et al.* 1985; Kvist *et al.* 1983). However, at least two other class I genes, *37* and *Mb1*, contain a similarly short intron 3 (Transy *et al.* 1987; Singer *et al.* 1988).

Exon 5 contains a splice-donor sequence that could splice the exon after 102 nucleotides to an exon 6. However, no sequences similar to a class I exon 6 are found in the 1253 bp of sequence that was determined beyond the splice-donor sequence. The *Thy19.4* cDNA clone isolated from the thymus cDNA library has unspliced and truncated 3' sequences, and thus does not reveal what 3' splicing occurs or where polyadenylation occurs in *Thy19.4* transcripts. Alternatively, exon 5 could encode the entire cytoplasmic domain and 3' untranslated tail seen in the cDNA clone. 237 nucleotides 3' of the putative termination codon is a 52 bp long A-rich tract followed by a 60 bp long, tandemly repeated, simple AAGG sequence. This tract is similar to polypyrimidine/polypurine asymmetrical tracts that are generally 200 bp long and are represented 10^5-10^6 times per genome (Birnboim *et al.* 1979). The function of these tracts is unknown, but there is evidence suggesting that they serve as transcriptional control elements (Drescher *et al.* 1987). There are two sites for polyadenylation (Proudfoot 1982), one 215 and the other 485 nucleotides 3' of the asymmetrical tract.

Putative Translation Product of Thy19.4. The putative product of this gene would be a 346 residue glycoprotein with several structural characteristics common

among classical class I molecules (Fig. 4). Translation would start in exon 1 with the initiator methionine, followed by a leader stretch of 31 amino acids. Exons 2, 3, and 4 would encode the approximately 90 amino acids in the external domains α1, α2, and α3. A 24 residue hydrophobic transmembrane domain, as well as an 11 residue cytoplasmic domain would be encoded by exon 5. The transmembrane domain is immediately followed by two positively charged lysine residues. Charged anchoring residues are typically found on the cytoplasmic side of transmembrane domains, and are proposed to prevent the short cytoplasmic domain from being pulled through the hydrophobic lipid bilayer (Warren 1981).

There are four potential N-linked glycosylation sites (Hubbard and Ivatt 1981): asparagines 86 and 90 in the α1 domain, asparagine 176 in the α2 domain, and asparagine 256 in the α3 domain. The first, third, and fourth potential sites are identical to those glycosylated in the $H\text{-}2K^d$ molecule (Kvist *et al.* 1983), but the second potential glycosylation site near the junction between the α1 and α2 domains is not found in other class I molecules. There are two cysteine residues in both the α2 and α3 domains, corresponding to those that form disulfide bonds within these domains in the other H-2 molecules. In addition, tryptophan 217 is at a position where that amino acid is invariably found in immunoglobulin supergene family domains. Interestingly, the putative Thy19.4 molecule would share an alanine to valine substitution at position 245 with human class I molecules HLA-Aw 68.1 and HLA-Aw 68.2. These two class I molecules do not associate with the CD8 molecule, because of the substitution at position 245, suggesting that alanine 245 is critical for CD8 binding (Salter *et al.* 1989). Since the association of CD8 with classical class I molecules is involved with their role in antigen presentation to T cells, the possible inability of the putative Thy19.4 molecule to associate with CD8 is consistent with its having a role other than antigen presentation to T

cells.

In addition to the extensive amino acid sequence conservation in the α3 domain (Table 1), other amino acids involved in the formation of the unique three-dimensional structure of the α1 and α2 domain of the HLA-A2 molecule are conserved in the putative Thy19.4 translation product (Bjorkman *et al.* 1987a, b; P. Bjorkman, personal communication). Proline 15 and glycine 16, which form a tight turn between β strands 1 and 2 in the HLA-A2 molecule, are found in the Thy19.4 translation. Proline 50 at a turn before the first α helix of α1 in the HLA-A2 molecule is conserved. In addition, glutamine 115, which contacts β_2-microglobulin in the HLA-A2 molecule, is also present. Many downward-pointing residues on the two α helices that contact the β-pleated sheet platform in the HLA-A2 molecule exhibit in the Thy19.4 amino acid sequence either identity or conservative substitutions. These include tryptophan 60, leucines 160, 168, 179 and alanine 153, as well as a conservative leucine-to-phenylalanine substitution at position 78. Finally, two salt bridges, histidine 3 with aspartic acid 29 and arginine 111 with aspartic acid 129, present in the HLA-A2 molecule and hypothesized to be present in other class I as well as in class II MHC molecules (Brown *et al.* 1988), are identical or conservatively substituted. The conservation of these critical residues suggests that the putative Thy19.4 molecule may have a classical class I molecule-like structure, and possibly may also function as a cell-surface recognition structure.

Expression of Thy19.4 RNA. To determine the extent of expression of the *Thy19.4* gene, PCR amplification was performed on single-stranded cDNA derived from 13 tissue- and developmental-stage RNA preparations (brain, heart, kidney, liver, lung, Peyer's patch, spleen, testis, and thymus, as well as 12-day, 16-day, 18-day

old fetus, and newborn). In addition, this analysis included single-stranded cDNA synthesized from total RNA from seven BALB/c mouse cell lines: WEHI-3 (myelomonocyte), I-10 (testicular Leydig cell), BALB/c CL.7 (fibroblast), BNL CL.2 (embryonic liver), J558 (myeloma), J774 A.1 (monocyte-macrophage), and BALB/3T3 (embryonic fibroblast). Primers from different exons were used to distinguish between amplifications of spliced mature mRNA, and unspliced nuclear RNA. Primers from exon 1 and 3 were used to determine the extent of expression of the *Thy19.4* gene, as well as whether these exons are properly spliced in *Thy19.4* transcripts. The predicted size of spliced cDNA that is amplified using the first and third exon primers is 456 bp, while amplified genomic DNA or unspliced cDNA is predicted to be 865 bp. Spliced message is amplified from all tissue cDNA preparations, except brain cDNA, where unspliced message is amplified (Fig. 5a). Since the brain RNA preparation was pretreated with RNAse-free DNAse prior to cDNA synthesis, it is most probable that the *Thy19.4* gene is transcribed, but not spliced, in the brain. Since unspliced *Thy19.4* transcripts can not be translated, it is unlikely that class I molecules encoded by the *Thy19.4* gene are expressed in that tissue. Similarly, classical class I molecules are not found in the brain. More intense bands are evident for amplified spleen, Peyer's patch, and thymus cDNA than for other tissues, suggesting that *Thy19.4* gene expression is highest in lymphoid tissues. Northern blots probed with a 5' *Thy19.4* probe detected a very low level, 2000 to 2200 nucleotide message in the thymus (data not shown). Thus, although *Thy19.4* gene is expressed at a low level, its pattern of expression is similar to that of the classical class I genes.

Thy19.4 gene expression is detected in all BALB/c tissue culture cells except J558, a myeloma cell line, and BALB 3T3, an embryonal fibroblast line (Fig. 5a). On the other hand, PCR amplification using *H-2D*^{d} oligonucleotides could detect

transcription of *H-2D^d* in J558 and BALB 3T3 (data not shown). Although at a low level relative to hematopoietic cell lines, *Thy19.4* gene expression is detected in I-10, a testicular cell line, and BALB/c CL.7, a fibroblast cell line. Since these cell lines are not hematopoietic in origin, the *Thy19.4* gene appears to be expressed in a wide range of cell types instead of merely infiltrating lymphocytes.

Since this assay is performed using *Thy19.4* gene specific oligonucleotides with an annealing temperature of 60°C, and the *Thy19.4* gene specific 5' probe, it is unlikely that transcripts from other class I genes are amplified. However, to test this possibility, cosmid clones containing the *H-2K^d*, *D^d*, *L^d*, *Q7^d*, *T13^c*, and *37^c* genes were amplified using the *Thy19.4* gene specific exon 1 and 3 oligonucleotides. No amplification is detected, using the *Thy19.4* 5' probe, from any of the cosmid clones that does not contain the *Thy19.4* gene (data not shown). Therefore, this assay probably specifically detects *Thy19.4* transcripts.

Genomic DNA and thymus cDNA were amplified using exon 3 and 5 primers to determine whether introns 3 and 4 are spliced from a mature *Thy19.4* transcript. The predicted size of amplified spliced cDNA using these two primers is 497 bp, while the predicted size of amplified genomic DNA is 1344 bp. Since a 497 bp fragment is amplified from thymus cDNA (Fig. 5b), it appears that introns 3 and 4 are removed from a mature *Thy19.4* transcript.

To determine the 3' splicing pattern and the polyadenylation signal used, thymus cDNA and genomic DNA were amplified using a 5th exon primer and two primers derived immediately 5' of the two polyadenylation signals. Both polyadenylation signal primers can, in conjunction with the fifth exon primer, amplify thymus cDNA (Fig. 5c). The fragment sizes produced by the amplification reactions are the same for the thymus cDNA and genomic DNA preparations. Taken together, these two results suggest that *Thy19.4* transcripts can extend at

least to the second polyadenylation signal, and that no splicing occurs 3' of exon 5. Northern blot data suggest that *Thy19.4* transcripts are approximately 2000 to 2200 nucleotides in length (data not shown). Assuming that 100 to 200 bases of polyadenosine is added to *Thy19.4* transcripts, the predicted size of transcripts using the second polyA signal is 2010 to 2110 nucleotides. Since no canonical polyA signal is found in the 375 bp sequenced beyond the second signal sequence, it is likely that the second signal is used to polyadenylate *Thy19.4* transcripts.

Mapping the Thy19.4 Gene. Restriction enzyme site polymorphisms and two recombinant mouse strains, R4-early (e) and R4-late (l), were used to map the *Thy19.4* gene (Fischer Lindahl 1986, Richards *et al.* 1989). The two strains have recombination breakpoints on either side of the *Hmt* gene locus, a class I gene that maps distal to the *H-2* and *Tla* regions of chromosome 17 (Fig. 6a). The two recombinants share the Hmt^b region with the *Mus musculus castaneus* strain CAS3, but the $H\text{-}2^k$, Qa^b, and Tla^b regions with the *Mus musculus domesticus* strain C3H (Fig. 6a). The strain R4-1 was generated by backcrossing R4-e with C3H.SW, and has a recombination breakpoint that maps between the *Hmt* locus and the more distal *Tpx-1* gene that encodes a testicular protein (Kasahara *et al.* 1987). The *Thy19.4* 3' probe detects the same-sized PstI fragment in the DNA from all three mice with CAS3-derived *Hmt* loci. A bigger fragment is seen in DNA from C3H and C3H.SW (Fig. 6b); therefore, the *Thy19.4* gene maps between the recombinational breakpoints in "R4-early" and "R4-late," inside the *Hmt* region on chromosome 17.

Discussion

This report presents the sequence of a novel non-classical BALB/cJ class I MHC

gene. Although structurally it resembles a functional class I gene, preserving open reading frames in all five exons and appropriate adjacent splice signals, the *Thy19.4* gene is fairly divergent in sequence from other BALB/c class I genes (Table 1). In the generally least-conserved leader and transmembrane exons (1 and 5), the sequence similarity between the *Thy19.4* gene and the $H\text{-}2K^d$, D^d, L^d, $Q7^d$, and $T13^c$ genes is between 51% and 58%. Exons 2 and 3 encoding the α1 and α2 domains are between 62% and 70% similar. Exon 4, which encodes the generally conserved α3 domain, displays 86% to 88% similarity to the others. In contrast, the $H\text{-}2K^d$ and $H\text{-}2D^d$ genes are 82% to 93% similar to each other in exons 1 through 5.

The *Thy19.4* gene is about equally similar to the $H\text{-}2K^d$, D^d, L^d, $Q7^d$, and $T13^c$ genes, suggesting that it had duplicated and started to diverge from these class I genes at about the same time in evolution as the most divergent of these, the $T13^c$ gene (Fisher *et al.* 1985). The *Thy19.4* gene is more dissimilar to the *Mb-1* gene than to the other class I genes. However, among mouse class I genes sequenced to date, the *Mb-1* gene is the most divergent (Singer *et al.* 1988). A dot-matrix plot of the *Thy19.4* gene versus the $H\text{-}2K^d$ gene (Fig. 7) shows that the coding regions of the *Thy19.4* gene are more conserved than intron sequences. This is consistent with the hypothesis that selection is acting on the exons of the *Thy19.4* gene to conserve its class I-like structure. Interestingly, the untranslated regions 5' of exon 1 and 3' of exon 5 have no detectable similarity. These regions were either not included in the duplication event that created the *Thy19.4* gene, or they diverged more rapidly than both coding region and intron sequences.

In spite of the nucleotide sequence divergence, the translation of the *Thy19.4* gene has conserved critical amino acids, which in the HLA-A2 molecule are involved in the formation of the three-dimensional structure. If this allows

the putative Thy19.4 product to have the unique tertiary structure of a classical class I molecule, and if that structure is important for function as a cell-surface recognition structure, then it is possible that the Thy19.4 molecule may also serve such a function. It is interesting to note that an interferon regulatory sequence is present in the 5' sequences of the *Thy19.4* gene, suggesting that transcription of this gene can respond to interferon induction, a notion possibly consistent with a functional role for this gene in the immune response. This role has been suggested for other transcriptionally active non-classical class I genes that map distal to the *H-2* complex. These genes include ubiquitously expressed genes like the *Thy19.4* and *37* genes (Transy *et al.* 1987), as well as tissue-specific genes like the thymus-specific $T13^c$ gene (Fisher *et al.* 1985), and the liver-specific *Q10* gene (Cosman *et al.* 1982). Similarly, their gene products have retained structural characteristics common among transplantation antigens, suggesting that they, like the transplantation antigens, may be involved in cell-cell interaction events. It is proposed that some of them may be differentiation antigens (Rothenberg 1982; Warner *et al.* 1987; Lynes *et al.* 1982), or serve as alternative antigen presenting structures (Fischer Lindahl *et al.* 1988). However, unlike any of the other previously characterized class I genes, the *Thy19.4* gene maps to a new MHC location, the *Hmt* region.

The *Hmt* region extends from the distal side of the *Tla* locus to a region proximal to the *Tpx-1* gene (Kasahara *et al.* 1989). The size of the *Hmt* region is between 0.1 and 1 cM, since both of these loci are 1-2 cM distal to the *H-2D* loci (Passmore and Romano 1988). Interestingly, the *Hmt* region also contains the *Pgk-2* gene, a testis-specific phosphoglycerate kinase gene (Boer *et al.* 1987; Richards *et al.* 1989), and the *Hmt* gene, a gene encoding a non-classical I molecule that associates with a mitochondrial peptide to form the maternally

transmitted antigen (Mta) (Fischer Lindahl *et al.* 1983, 1988). Since both the *Thy19.4* and *Hmt* genes map to this region, it is clear that there are more class I genes than originally found by cosmid cloning (Steinmetz *et al.* 1982).

The finding of L1 repeats surrounding the *Thy19.4* gene suggests that the *Hmt* region may reside in a different type of chromatin than the *H-2* or *Tla* regions. Long and short interspersed repetitive elements (LINES and SINES) are reported to be distributed unevenly throughout the human genome, mapping to areas differentiated by chromosomal staining patterns (Korenberg and Rykowski 1988). SINES include the Alu-like repetitive elements that are present throughout the *H-2* and *Tla* region of the MHC (Krayev *et al.* 1980, 1982, Moore *et al.* 1982, Fisher *et al.* 1985). They are relatively clustered in the reverse staining bands of the chromosomes, areas of high transcriptional activity. On the other hand, LINES, which include the L1 repeat elements at both 5' and 3' of the *Thy19.4* gene (Fig. 2; Singer and Skowronski 1985), are concentrated in the Giemsa staining that bands of the chromosomes, areas of low transcriptional activity (Korenberg *et al.* 1978; Goldman *et al.* 1984). It is possible that the low level of expression of the *Thy19.4* gene is a reflection of the low transcriptional activity of the chromatin in which it resides.

Acknowledgements

Supported by NIH grant AI-17565. The authors wish to thank Drs. P. Bjorkman, I. Stroynowski, M. Zuniga, D. Nickerson, and J. Kobori for critically reviewing this manuscript and Mrs. C. Blagg for expert secretarial assistance.

References

Aaronson, S., and Todaro, G. Development of 3T3-like lines from BALB/c mouse embryo cultures: Transformation susceptibility to SV40. *J. Cell. Physiol. 72*: 141-148, 1968.

Birnboim, H., Sederoff, R., and Paterson, M. Distribution of polypyrimidine/polypurine segments in DNA from diverse organisms. *Eur. J. Biochem. 98:* 301-307, 1979.

Bjorkman, P., Saper, M., Samraoui, B., Bennett, W., Strominger, J., and Wiley, D. Structure of the human class I histocompatibility antigen, HLA-A2. *Nature 329*: 506-512, 1987.

Bjorkman, P., Saper, M., Samraoui, B., Bennett, W., Strominger, J., and Wiley, D. The foreign antigen-binding site and T-cell recognition regions of class I histocompatibility antigens. *Nature 329*: 512-518, 1987.

Boer, P., Adra, C., Lau, Y., and McBurney, M. The testis-specific phosphoglycerate kinase gene *Pgk-2* is a recruited retroposon. *Mol. Cell. Biol. 7:* 3107-3112, 1987.

Breathnach, R., and Chambon, P. Organization and expression of eukaryotic split genes coding for proteins. *Ann. Rev. Biochem. 50:* 349-383, 1981.

Brown, J. H., Jardetzky, T., Saper, M., Samraoui, B., Bjorkman, P., and Wiley, D. A hypothetical model of the foreign antigen-binding site of class II histocompatibility molecules. *Nature 332*: 845-850, 1988.

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

Drescher, U., Chowdhury, K., and Gruss, P. Isolation and characterizaiton of murine transcriptional control elements using a "shotgun" method. *DNA 6:* 307-316, 1987.

Feinberg, A. P., and Vogelstein, B. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem. 132:* 6-13, 1983.

Fischer Lindahl, K. Genetic variants of histocompatibility genes from wild mice. *Curr. Top. Microbiol. Immunol. 127:* 272-278, 1986.

Fischer Lindahl, K., Hausmann, B., and Chapman, V. M. A new *H-2*-linked class I gene whose expression depends on a maternally inherited factor. *Nature 306:* 383-385, 1983.

Fischer Lindahl, K., Loveland, B. E., and Richards, C. S. The end of *H-2*. *In* C. S. David (ed.). *Major Histocompatibility Genes and Their Role in Immune Function* , Plenum Press, New York, 1988, pp. 327-338.

Fisher, D. A., Hunt, S. W., and Hood, L. 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, 1985.

Flaherty, L. The *Tla* region of the mouse: Identification of a new serologically defined locus, *Qa-2*. *Immunogenetics 3:* 533-539, 1976.

Goldman, M., Holmquist, G., Gray, M., Caston, L., and Nagy, A. Replication timing of genes and middle repetitive sequences. *Science 224:* 686-692, 1984.

Hubbard, S., and Ivatt, R. Synthesis and processing of asparagine-linked oligosaccharides. *Ann. Rev. Biochem. 50:* 555-583, 1981.

Hutchison, C., Hardies, S., Padgett, R., Weaver, S., and Edgell, M. The mouse globin pseudogene *bh3* is descended from a premammalian δ-*globin* gene. *J. Biol. Chem. 259:* 12881-12889, 1984.

Huynh, T., Young, R., and Davis, R. Constructing and screening cDNA libraries in λgt10 and λgt11. *In* D. Glover (ed.) *DNA Cloning: A Practical Approach,* IRL Press, Oxford, 1984, pp. 49-78.

Israel, A., Kimura, A., Fournier, A., Fellous, M., and Kourilsky, P. Interferon response-sequence potentiates activity of an enhancer in the promoter region of a mouse *H-2* gene. *Nature 322:* 743-746, 1986.

Kasahara, M., Figueroa, F., and Klein, J. Random cloning of genes from mouse chromosome 17. *Proc. Natl. Acad. Sci. USA 84:* 3325-3328, 1987.

Kasahara, M., Passmore, H., and Klein, J. A novel testis-specific gene *Tpx-1* maps between *Pgk-2* and *Mep-1* on mouse chromosome 17. *Immunogenetics 16:* 319-328, 1989.

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

Korber, B., Mermod, N., Hood, L., and Stroynowski, I. Regulation of gene expression by interferons: Control of *H-2* promoter responses. *Science 239:* 1302-1306, 1988.

Korenberg, J., Therman, E., and Denniston, C. Hot spots and functional organization of human chromosomes. *Hum. Genet. 43:* 13-22, 1978.

Korenberg, J. R. and Rykowski, M. C. Human gene organization: Alu, Lines, and the molecular structure of metaphase chromosome bands. *Cell 53:* 391-400, 1988.

Kozak, M. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucl. Acids Res. 12:* 857-872, 1984.

Krayev, A., Kramerov, D., Skryabin, K., Ryskov, A., Bayev, A., and Georgiev, G. The nucleotide sequence of the ubiquitous repetitive DNA sequence B1 complementary to the most abundant class of mouse fold-back RNA. *Nucl. Acids Res. 8:* 1201-1215, 1980.

Krayev, A., Markusheva, T., Kramerov, D., Ryskov, A., Skryabin, K., Bayev, A., and Georgiev, G. Ubiquitous transposon-like repeats B1 and B2 of the mouse genome: B2 sequencing. *Nucl. Acids Res. 10:* 7461-7475, 1982.

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

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

Lundblad, A., Steller, R., Kabat, E., Hirst, J., Weigert, M., and Cohn, M. Immunochemical studies on mouse myeloma proteins with specificity for dextran or levan. *Immunochemistry 9:* 535-544, 1972.

Lynes, M. A., Tonkonogy, S., and Flaherty, L. Qa-1 and Qa-2 expression on CFU-s. *J. Immunol. 129:* 928-930, 1982.

Maniatis, T., Fritsch, E. F., and Sambrook, J. Molecular Cloning. *In A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1982.

Metcalf, D., Moore, M., and Warner, N. Colony formation *in vitro* by myelomonocytic leukemia cells. *J. Natl Cancer Inst. 43:* 983-997, 1969.

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

Mount, S.. A catalogue of splice junction sequences. *Nucl. Acids Res. 10:* 459-472, 1982.

Old, L. J., Boyse, E. A., and Stockert, E. Antigenic properties of experimental leukemias. I. Serological studies *in vitro* with spontaneous and radiation-induced leukemias. *J. Natl. Cancer Inst. 31:* 977-986, 1963.

Passmore, H. and Romano, J. Genetic organization of the *Qa* and *Tla* regions: Gene mapping based on the analysis of recombinant strains. *In* C. S. David (ed.) *Major Histocompatibility Genes and Their Role in Immune Function,* Plenum Press, New York, 1988, pp. 49-60.

Patek, P., Collins, J., and Cohn, M. Transformed cell lines susceptible or resistant to *in vivo* surveillance against tumorigenesis. *Nature 176:* 510-511, 1978.

Proudfoot, N. The end of the message. *Nature 298:* 516-517, 1982.

Ralph, P., and Nakoinz, I. Inhibitory effects of lectins and lymphocyte mitogens on murine lymphomas and myelomas. *J. of the Natl. Cancer Inst. 51:* 883-890, 1973.

Richards, S., Bucan, M., Brorson, K., Kiefer, M., Hunt, S., Lehrach, H., and Fischer Lindahl, K. Genetic and molecular mapping of the *Hmt* region of the mouse. *EMBO J.*, 1989, in press.

Robinson, P. J. Qb-1, a new class I polypeptide encoded by the *Qa* region of the mouse *H-2* complex. *Immunogenetics 22:* 285-289, 1985.

Rothenberg, E. A specific biosynthetic marker for immature thymic lymphoblasts. Active synthesis of thymus-leukemia antigen restricted to proliferating cells. *J. Exp. Med 155:* 140-154, 1982.

Saiki, R., Gelfand, D., Stoffel, S., Scharf, S., Higuchi, R., Horn, G., Mullis, K., and Erlich, H. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science 239:* 487-494, 1988.

Salter, R. D., Norment, A. M., Chen, B., Clayberger, C., Krensky, A., Littman, D., and Parham, P. Polymorphism in the α3 domain of HLA-A molecules affects binding to CD8. *Nature 338*: 345-347, 1989.

Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H., and Roe, B. A. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol. 143:* 161-178, 1980.

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

Silver, J. and Hood, L. Detergent-solubilized H-2 alloantigen is associated with a small molecular-weight polypeptide. *Nature 249*: 764-765, 1974.

Singer, D. S., Hare, J., Golding, H., Flaherty, L., and Rudikoff, S. Characterization of a new subfamily of class I genes in the *H-2* complex of the mouse. *Immunogenetics 28:* 13-21, 1988.

Singer, M., and Skrowronski, J. Making sense out of LINES: long interspersed repeat sequences in mammalian genomes. *TIBS 10:* 119-122, 1985.

Stanton, T. H. and Boyse, E. A. A new serologically defined locus, *Qa-1*, in the *Tla* region of the mouse. *Immunogenetics 3:* 525-531, 1976.

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

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

Transy, C., Nash, S. R., David-Watine, B., Cochet, M., Hunt, S. W., Hood, L. E., and Kourilsky, P. A low polymorphic mouse *H-2* class I gene from the *Tla* complex is expressed in a broad variety of cell types. *J. Exp. Med. 166:* 341-362, 1987.

Warner, C. M., Gollnick, S. O., Flaherty, L., and Goldbard, S. B. Analysis of Qa-2 antigen expression by preimplantation mouse embryos: Possible relationship to the preimplantation embryo-development (*Ped*) gene product. *Biology of Reproduction 36:* 611-616, 1987.

Warren, G. Membrane proteins: Structure and assembly. *In* J. B. Finean and R. H. Michell (eds.) *Membrane Structure.* Elsevier/North Holland Biomedical Press, Amsterdam, Holland, 1981, pp. 215-217.

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

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

Yasumura, Y., Tashijian, A., and Sato, G. Establishment of four functional, clonal strains of animal cells in culture. *Science 154:* 1186-1189, 1966.

Zinkernagel, R. M., and Doherty, P. C. MHC-restricted cytotoxic T cells: studies on the role of polymorphic major transplantation antigens determining T-cell restriction specificity, function, and responsiveness. *Adv. Immunol. 27:* 51-177, 1980.

Table 1

Amino Acid and Nucleotide Identities

		K^d vs. D^d	Thy 19.4 vs. K^d	Thy 19.4 vs. D^d	Thy 19.4 vs. L^d	Thy 19.4 vs. Qa 2, 3(Q7)	Thy 19.4 vs. Tla^c (T13)	Thy 19.4 vs. Mb1 (C57BL/10)
Exon 1	Amino Acid	90%	24%	21%	20%	28%	12%	25%
	Nucleotide	86%	58%	53%	57%	54%	54%	52%
Exon 2	Amino Acid	79%	62%	64%	59%	57%	47%	31%
	Nucleotide	89%	67%	67%	68%	65%	62%	54%
Exon 3	Amino Acid	83%	63%	61%	61%	59%	55%	33%
	Nucleotide	91%	68%	68%	68%	69%	67%	56%
Exon 4	Amino Acid	86%	78%	78%	77%	76%	75%	64%
	Nucleotide	93%	88%	88%	88%	88%	86%	70%
Exon 5	Amino Acid	77%	26%	24%	21%	18%	12%	12%
	Nucleotide	82%	52%	52%	70%	54%	51%	62%

Comparison of *Thy19.4* nucleotide and amino acid sequences with six other class I genes. Nucleotide and amino acid sequence identities are shown for exons 1-5. For reference, the *H-2K*^{d} and *H-2D*^{d} genes are compared to each other as well.

Fig. 1. Map of class I genes in the BALB/c MHC. Class I genes map to the *K*, *D*, *Qa*, and *Tla* regions. The *I* and *S* regions contain class II and complement genes, respectively. The order of the *Tla* region gene clusters is unknown, as is the distance between the *K*, *D*, and *Tla* clusters. The *Thy19.4* gene maps to a new region distal to the *Tla* region, the *Hmt* region.

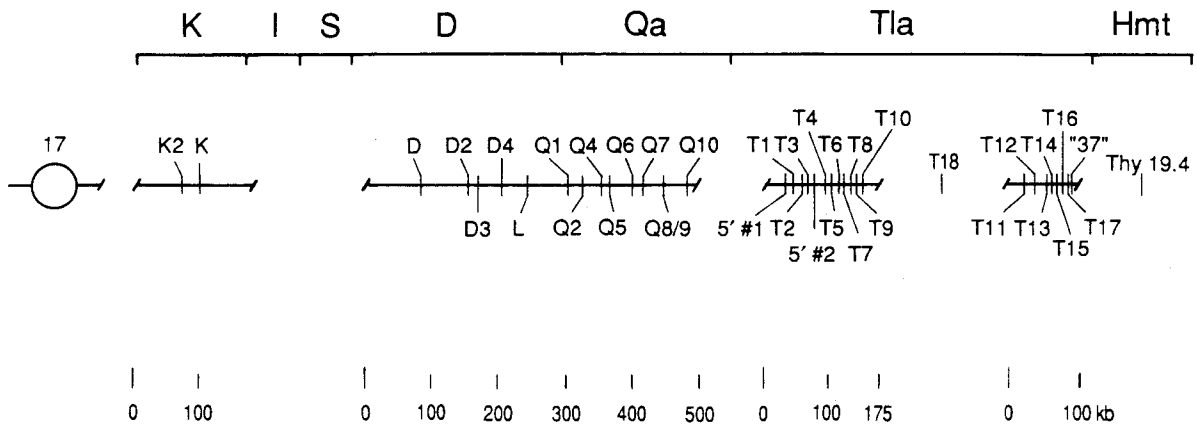

K
I
S
D
Qa
Tla
Hmt
17
K2 K
D D2 D4 Q1 Q4 Q6 Q7 Q10
D3 L Q2 Q5 Q8/9
T4 T10
T1 T3 T6 T8
5′ #1 T2 T5 T9
5′ #2 T7
T18
T16
T12 T14 "37"
T11 T13 T17
T15
Thy 19.4
0 100
0 100 200 300 400 500
0 100 175
0 100 kb

Fig. 2. Restriction map of the *Thy19.4* gene. Putative exons are denoted by boxes and L1 repeat elements by arrows. EcoRI sites are denoted by R, BamHI by B, HindIII by H, and BglII by G. The DNA between the 5' HindIII site and the third EcoRI site was sequenced completely on both strands. The 1200 bp 5' and the 550 bp 3' probes are shown as bars below the map.

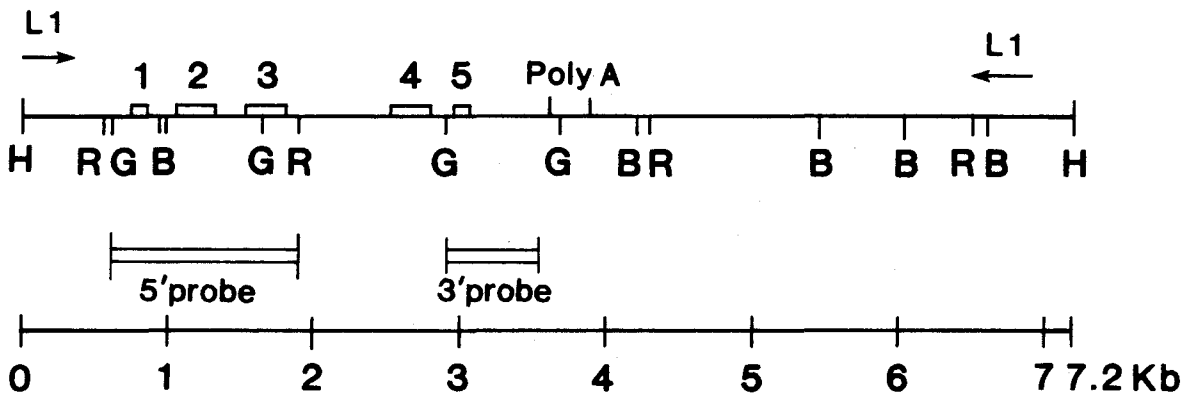

L1
1 2 3 4 5 Poly A
L1
H R G B G R G G B R B B R B H
5'probe
3'probe
0 1 2 3 4 5 6 7 7.2 Kb

Fig. 3. Southern blot of genomic DNA from a variety of wild mice digested with the restriction enzyme BamHI and probed with the *Thy19.4* 3' probe. Arrows point to faint bands.

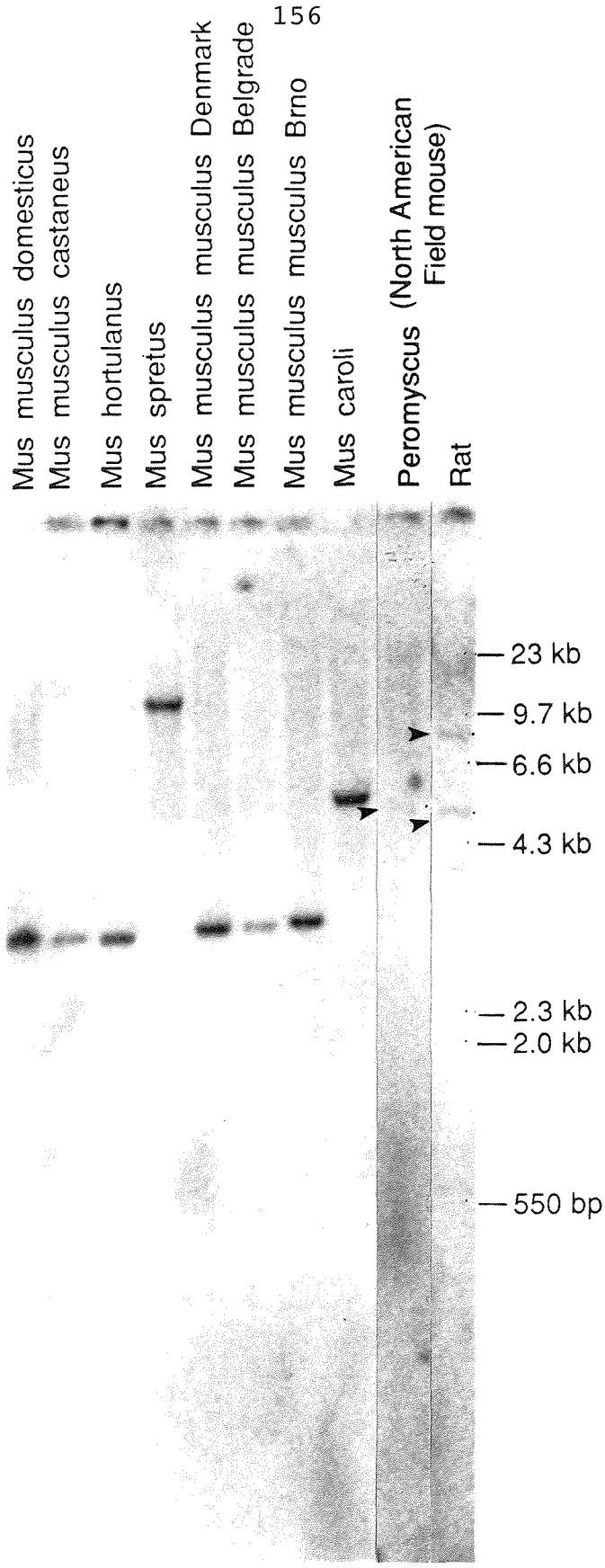

Mus musculus domesticus
Mus musculus castaneus
Mus hortulanus
Mus spretus
Mus musculus musculus Denmark
Mus musculus musculus Belgrade
Mus musculus musculus Brno
Mus caroli
Peromyscus (North American Field mouse)
Rat
23 kb
9.7 kb
6.6 kb
4.3 kb
2.3 kb
2.0 kb
550 bp

Fig. 4. DNA Sequence of the *Thy19.4* gene. At the 5'-end, the L1 repeat extends until bp 329. CAAT and TATA boxes as well as interferon regulatory sequences are underlined 5' of exon 1. The translation is shown for exon sequences. A donor splice signal in exon 5 is marked by arrows. The potential glycosylation sites in the α 1, 2, and 3 domains are boxed, while the conserved cysteine residues in the α2 and α3 domains are circled. The two poly(A) addition signals that are found 3' of the gene are underlined, as is the A-rich tract.

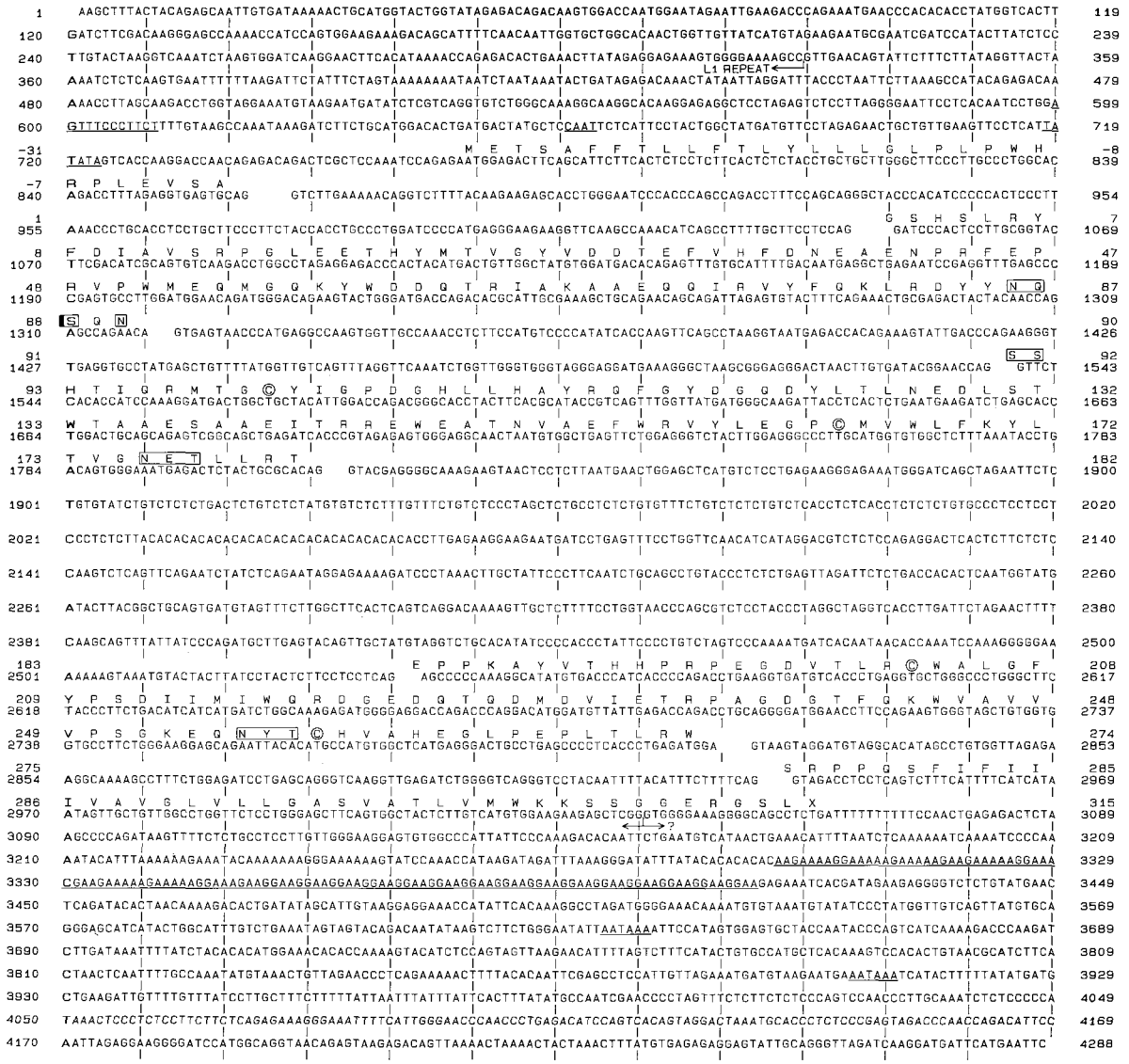
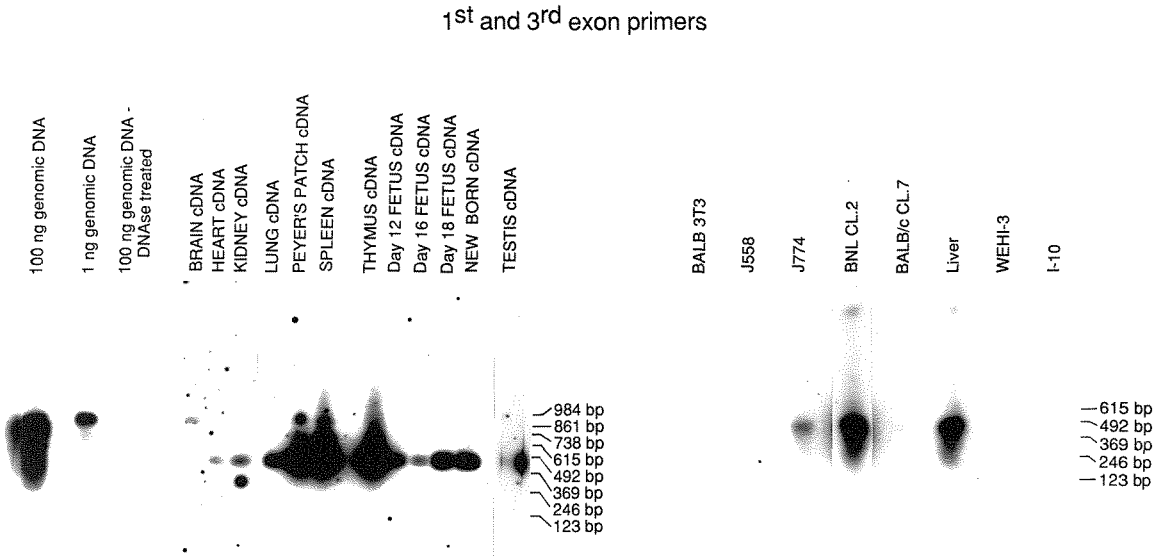

L1 REPEAT

Fig. 5a-c. (a) Southern blot of PCR amplified first strand cDNA or genomic DNA using exon 1 and 3 oligonucleotides. The blot was probed with a 5' probe generated by isolating a 1.3 kb EcoRI fragment containing the first, second, and third exons of the *Thy19.4* gene. Exposure times were: brain, heart, kidney, liver, testis, BALB 3T3, J558, J774A.1, BALB CL.7, I-10, WEHI-3 ~4 days; lung, Peyer's patch, spleen, thymus, feti, newborn, BNL CL.2 ~16 hours; genomic DNA ~1 hour. Spliced transcripts are detected in all tissues except the brain. In addition, all cell lines express the *Thy19.4* gene with the exception of BALB 3T3 and J558. The *Thy19.4* gene cannot be amplified from 100 ng of genomic DNA pretreated with RNAse-free DNAse, demonstrating that DNAse treatment removes any contaminating DNA that may have been present initially in the RNA preparations.

(b) Southern blot of PCR amplified thymus cDNA or genomic DNA using exon 3 and 5 oligonucleotides. The blot was probed with the 5' and 3' *Thy19.4* probes. Exposure times were ~1 hour for genomic DNA and ~16 hours for thymus cDNA. A 497 bp band is amplified from thymus cDNA, the correct-sized for a fully spliced transcript. An arrow points to the correct sized band amplified from genomic DNA. Smaller molecular weight bands result from secondary priming during the PCR reaction.

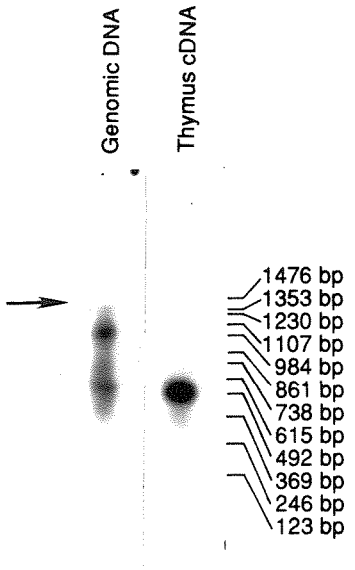
(c) Southern blot of PCR amplified thymus cDNA or genomic DNA using 3' oligonucleotides. The blot was probed with the 3' *Thy19.4* probe.

a


1st and 3rd exon primers
100 ng genomic DNA
1 ng genomic DNA
100 ng genomic DNA - DNAse treated
BRAIN cDNA
HEART cDNA
KIDNEY cDNA
LUNG cDNA
PEYER'S PATCH cDNA
SPLEEN cDNA
THYMUS cDNA
Day 12 FETUS cDNA
Day 16 FETUS cDNA
Day 18 FETUS cDNA
NEW BORN cDNA
TESTIS cDNA
BALB 3T3
J558
J774
BNL CL.2
BALB/c CL.7
Liver
WEHI-3
I-10
984 bp
861 bp
738 bp
615 bp
492 bp
369 bp
246 bp
123 bp
615 bp
492 bp
369 bp
246 bp
123 bp

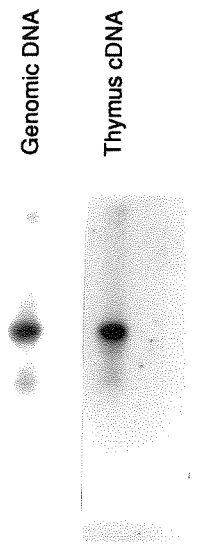
b

3rd and 5th exon primers



c

5th exon and 1st poly A signal primers



5th exon and 2nd poly A signal primers

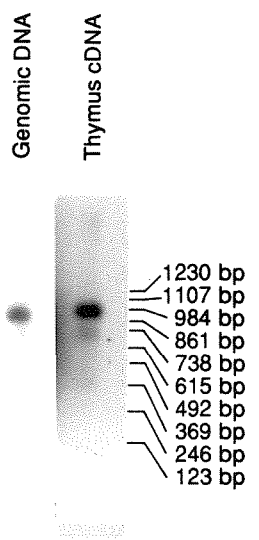
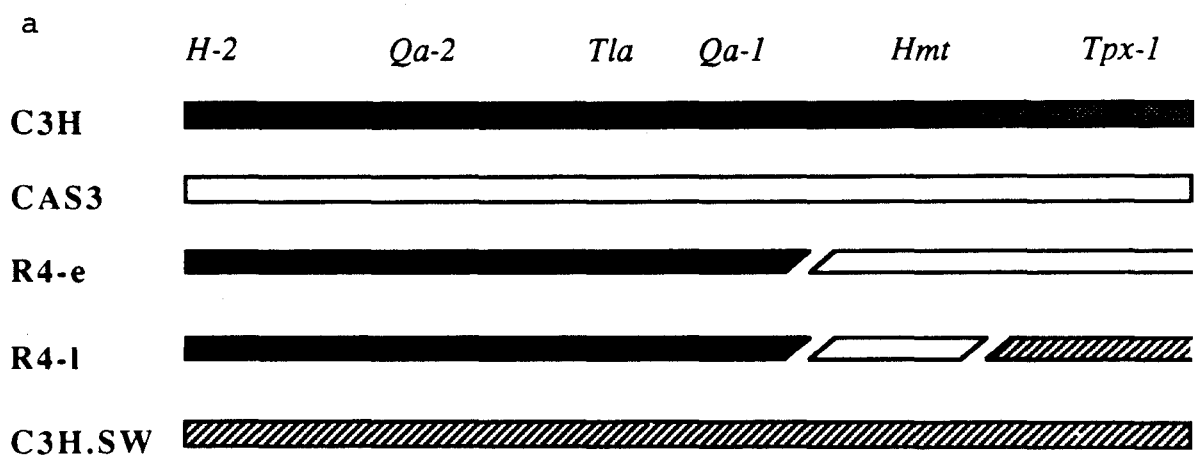


Fig. 6a-b. ***Mapping of Thy19.4 to the Hmt region.***

(a) The MHC haplotypes of the R4-e and CSW.R4-1 recombinant mice and their parental strains. R4-e is a recombinant between C3H/HeJ (C3H) and *M. m. castaneus* (CAS3) mice. During backcrossing of R4-e to C3H.SW, a second recombination occurred, yielding the R4-1 recombinant mouse. The origin of the *H-2*, *Qa-2*, and *Tla* alleles in the recombinant mice was determined by serotyping, while the origin of the *Qa-1*, and *Hmt* alleles was ascertained by CTL typing. The origin of the *Tpx-1* loci was determined by Southern blotting (Richards *et al.* 1989).

(b) PstI Southern blot of C3H, C3H.SW, R4-e, CSW.R4-1 and CAS3 DNA probed with the *Thy19.4* 3' probe.

a



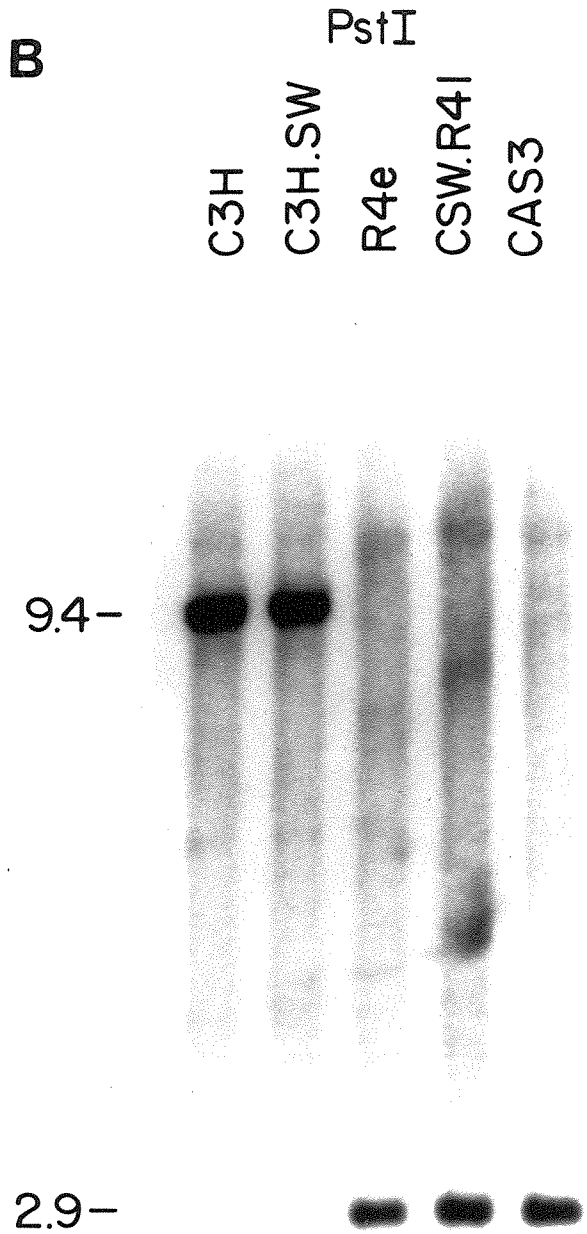
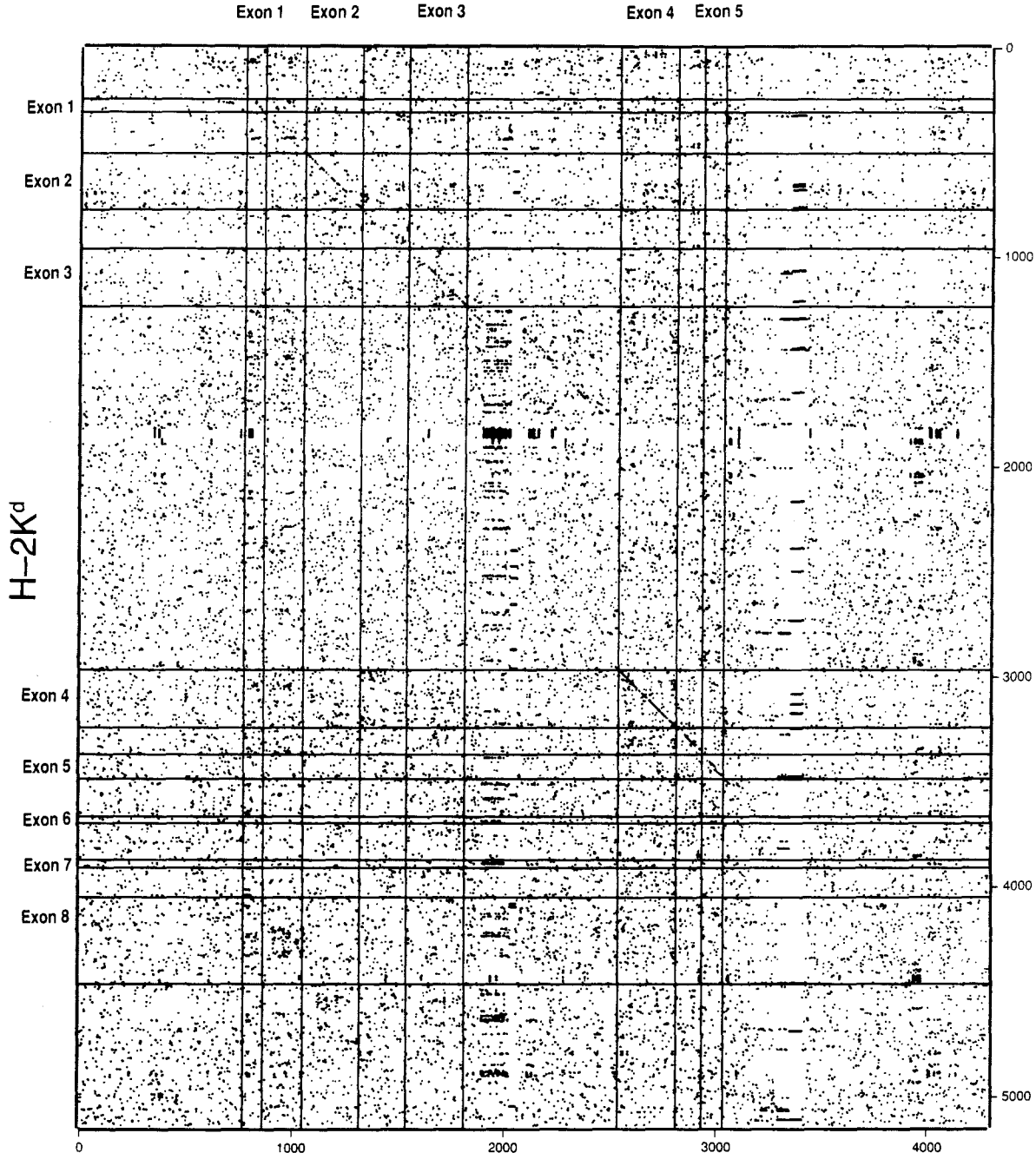

B
PstI
C3H
C3H.SW
R4e
CSW.R41
CAS3
9.4–
2.9–

Fig. 7. Dot-matrix analysis of the *Thy19.4* and *H-2K*^{d} genes. The exon and intron structures are shown on the sides of the figure and are boxed within the graph. Each dot represents a match of 6 of 7 nucleotides between the two sequences. Lines of homology are evident between both intron and exon sequences, but are most evident between exons. Sequences of repeated NT, where N is any nucleotide, between nucleotides 1893-1948 and repeated AAGG between nucleotides 3348 and 3411 are evident as a series of horizontal lines.

Thy 19.4


Exon 1
Exon 2
Exon 3
Exon 4
Exon 5
Exon 1
Exon 2
Exon 3
Exon 4
Exon 5
Exon 6
Exon 7
Exon 8
H-2K^d
0
1000
2000
3000
4000
5000
0
1000
2000
3000
4000

CONCLUSION

The major histocompatibility complex (MHC) of the BALB/c mouse contains at least 35 class I genes. These genes include three that encode the transplantation antigens, K^d, D^d, and L^d, as well as others that encode non-classical class I molecules that structurally resemble the transplantation antigens, but have no known function. However, the majority of these genes have not been characterized beyond hybridization experiments that mapped them on a series of MHC cosmids. The aim of this thesis was to initiate an examination of the family of non-classical class I genes by identifying genes that are expressed, determining whether they have open reading frames, and possibly interpreting functional aspects given their structure and evolution.

In Chapter 1, the DNA sequence of the 35 class I genes from the BALB/c mouse is analyzed in one segment, the transmembrane encoding fifth exon. On the basis of their nucleotide sequence, these sequences can be assigned to seven non-overlapping groups. The individual members of the different groups share little nucleotide sequence similarity, and hence are distantly related evolutionarily. The transmembrane domains could have diverged because they are selected only to maintain hydrophobicity, or alternatively because they are adapted to new functions. However, in spite of their extensive variation, the majority of the fifth exons appear to resemble those of the transplantation antigen genes in that they encode a proline-rich connecting peptide, a hydrophobic transmembrane segment, and a cytoplasmic domain with basic anchoring residues. Interestingly, several of the putative transmembrane domains encoded by divergent *Tla* region class I genes appear to be sufficiently amphipathic to promote interactions within the membrane with other proteins.

The second chapter describes 15 class I transcripts that are expressed in the adult thymus. By constructing and analyzing a thymus cDNA library, it was

demonstrated that a wide range of class I genes are expressed in the thymus, including several divergent genes from the *Tla* and *Hmt* regions. These non-classical class I genes display high nucleotide sequence similarity (>80%) to the classical class I genes in exon 4, which encodes the external α3 domain, intermediate similarity (50-70%) in exons 2 and 3, and little significant similarity (<50%) in any of the remaining portions. The presence of a wide range of transcripts from divergent non-classical class I genes in the thymus suggests that class I cell-surface recognition structures that are not restriction elements are expressed and perhaps functional in this tissue.

Chapter 3 is an in depth study of a novel class I gene, *Thy19.4*. This class I gene contains an open reading frame, and in several ways resembles the genes that encode the transplantation antigens. These similarities include shared exon/intron structure, and expression in a variety of tissues. In addition, its deduced gene product has conserved certain amino acids in the external domains (α1, α2, and α3), which in the human HLA-A2 class I molecule appear to be important for the formation of tertiary structure (Bjorkman *et al.*, 1987; P. Bjorkman, personal communication). However, the putative molecule encoded by this gene is divergent from transplantation antigens at the amino acid sequence level in the α1, α2, transmembrane, and cytoplasmic domains. Thus, although it is possible that this molecule will fold and associate with β_2-microglobulin in a manner similar to the transplantation antigens, it is unlikely that it has identical external recognition and internal cytoplasmic functions.

Significance of Findings. Histocompatibility systems are evolutionarily ancient and have been observed in several multicellular organisms, including angiosperm plants (East and Mangelsdorf 1925); cnidarians (Theodor 1970); annelids (Cooper 1969); and colonial marine protochordates (Bancroft 1903; Oka

and Watanabe 1957). Since the vertebrates probably evolved from marine protochordates (Young 1962), it is possible that the present vertebrate MHC antigens shared a common ancestor with their histocompatibility systems (Scofield *et al.* 1982). When the vertebrates evolved from the protochordates, this system was retained and became part of the immune system. In addition to the immune function, it is suggested that other descendants of the primitive histocompatibility molecules could perform other functions as cell surface recognition units, such as being targets for tissue necrosis during development (Williams 1982). However, it is unclear whether any of the putative divergent non-classical class I molecules serve this function. At least two, the *T18*^{c} and *Thy19.4* genes, appear to predate bird/mammal divergence (300 million years ago) and contain open reading frames. Hence, they have been selected for in the vertebrate genome for an extensive period of evolutionary history (Figs. 1 and 2a). If they first appeared in an early vertebrate, it is possible that they may have primordial recognition functions, such as that required by development, common to several types of vertebrates.

The evolution and structure of the non-classical class I genes provide some hints about their possible functions. Although the *Tla* and *Hmt* region class I genes have overall diverged significantly from the transplantation antigen genes, their α3 domains, encoded by the fourth exons, and to a lesser extent their α1 and α2 domains, encoded by the second and third exons, share similarity. The *Thy19.4* gene, for example, shares limited nucleotide sequence similarity with the transplantation antigen genes in the second and third exon. However, the α1 and α2 domains of the putative Thy19.4 molecule has conserved critical structural amino acids, which suggest that its structure may be similar to that of the transplantation antigens. Interestingly, the analysis of silent and replacement

substitutions reveals that the *Thy19.4* gene has more selective pressure exerted on its second and third exon coding sequences than the other non-classical class I genes (Table I). It is conceivable that this selective pressure is exerted on the second and third exons of the *Thy19.4* gene to conserve coding sequences that permit the putative Thy19.4 molecule to assume a transplantation antigen-like structure in the α1 and α2 domains. Among all of the non-classical class I genes, the α3 domain-encoding fourth exon is the most conserved. Interestingly, this appears to have resulted primarily from gene conversion events (Hayashida and Miyata 1983; Fisher *et al.* 1985; Figs. 2a and b). Since their α3 domains are conserved, it is possible that the putative non-classical class I molecules could associate with β_2-microglobulin. Finally, the majority of the fifth exons can encode domains that are sufficiently hydrophobic to anchor a protein in a lipid bilayer. However, the transmembrane domains encoded by the divergent *Tla* region genes are completely dissimilar to those of the transplantation antigens, and thus probably have different functions. The potential amphipathicity of several of the *Tla* region transmembranes suggests that they, unlike those of the restriction elements, could associate within the membrane with other proteins, possibly to initiate signaling cascades or trafficking.

Therefore, it is unlikely that the recognition functions of the external domains or the internal functions of the cytoplasmic domains are identical to those of the transplantation antigens. Instead, it is likely that they are β_2-microglobulin-binding, cell-surface recognition structures that perform functions other than antigen presentation. In at least one tissue, the adult thymus, a wide range of non-classical class I genes are expressed, possibly because they are involved in its function as the site of T-cell differentiation. Whatever their function is, their conservation appears to be important since they have been

maintained in the vertebrate genome since the Paleozoic era 300 million years ago.

References

Bancroft, R. Variation and fusion of colonies in compound ascidians. *Proc. Calif. Acad. Sci.* **3**:137-187, 1903.

Bjorkman, P., Saper, M., Samraoui, B., Bennett, W., Strominger, J., and Wiley, D. Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* **329**:506-512, 1987.

Cooper, E. Chronic allograft rejection in *Lumbricus terrestris*. *J. Exp. Zool.* **171**:69-74, 1969.

East, E., and Mangelsdorf, A. A new interpretation of the hereditary behavior of self-sterile plants. *Proc. Natl. Acad. Sci. USA* **11**:166-171, 1925.

Fisher, D., Hunt, S., and Hood, L. 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**:528-545, 1985.

Hayashida, H., and Miyata, T. Unusual evolutionary conservation and frequent DNA segment exchange in class I genes of the major histocompatibility complex. *Proc. Natl. Acad. Sci. USA* **80**:2671-2675, 1983.

Li, W., Wu, C., and Luo, C. A new method for estimating synonymous and non-synonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon charges. *Mol. Biol. Evol.* **2**:150-174, 1985.

Lillegraven, J., Kielan-Laworowska, Z., and Clemens, W. *Mesozoic Mammals, the First Two-Thirds of Mammalian History*. University of California Press, Berkeley, 1979.

Nei, M. *Molecular Evolutionary Genetics*. Columbia University Press, New York, 1987.

Oka, H., and Watanabe, H. Colony-specificity in compound ascidians as tested by fusion experiments. *Proc. Japan Acad.* **33**:657-659, 1957.

Scofield, V., Schlumpberger, J., West, L., and Weissman, I. Protochordate allorecognition is controlled by a MHC-like gene system. *Nature* **295**:499-502, 1982.

Theodor, J. Distinction between "self" and "not-self" in lower invertebrates. *Nature* **227**:690-692, 1970.

Williams, A. Surface molecules and cell interactions. *J. Theor. Biol.* **98**:221-231, 1982.

Young, J. *The Life of Vertebrates.* Oxford University Press, New York, 1962.

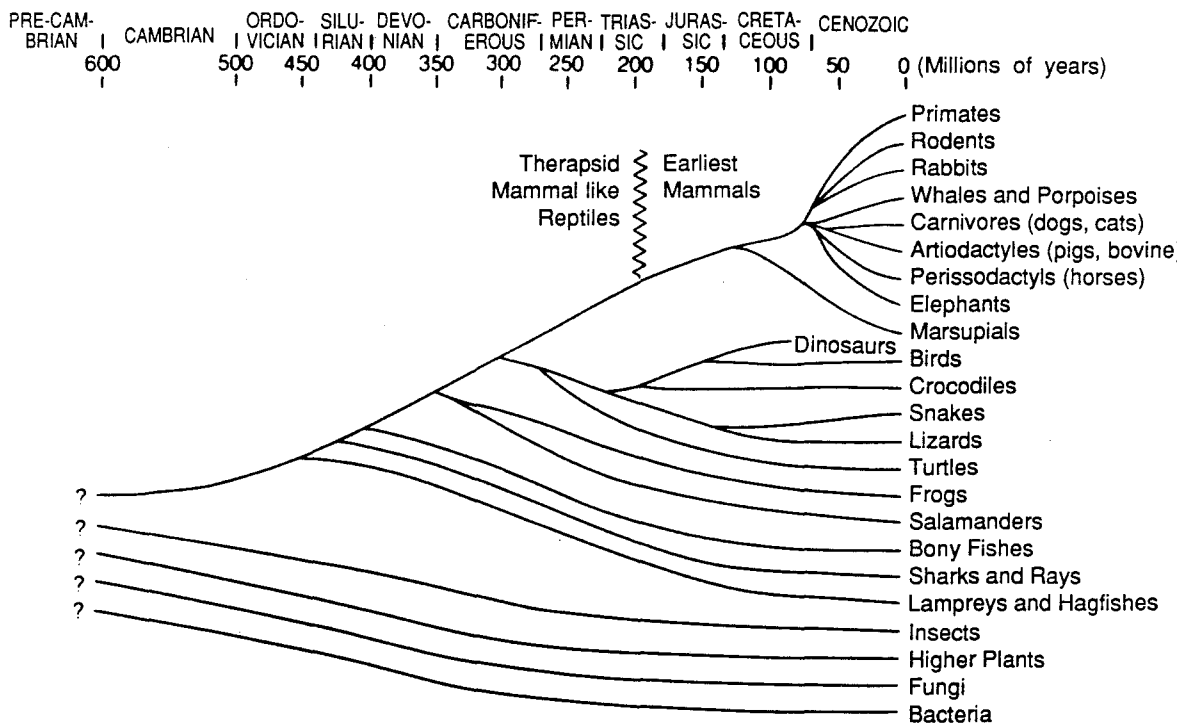
Table I

Comparison of Silent-to-Replacement Frequency Ratios Among Various Mammalian Genes and Class I Genes

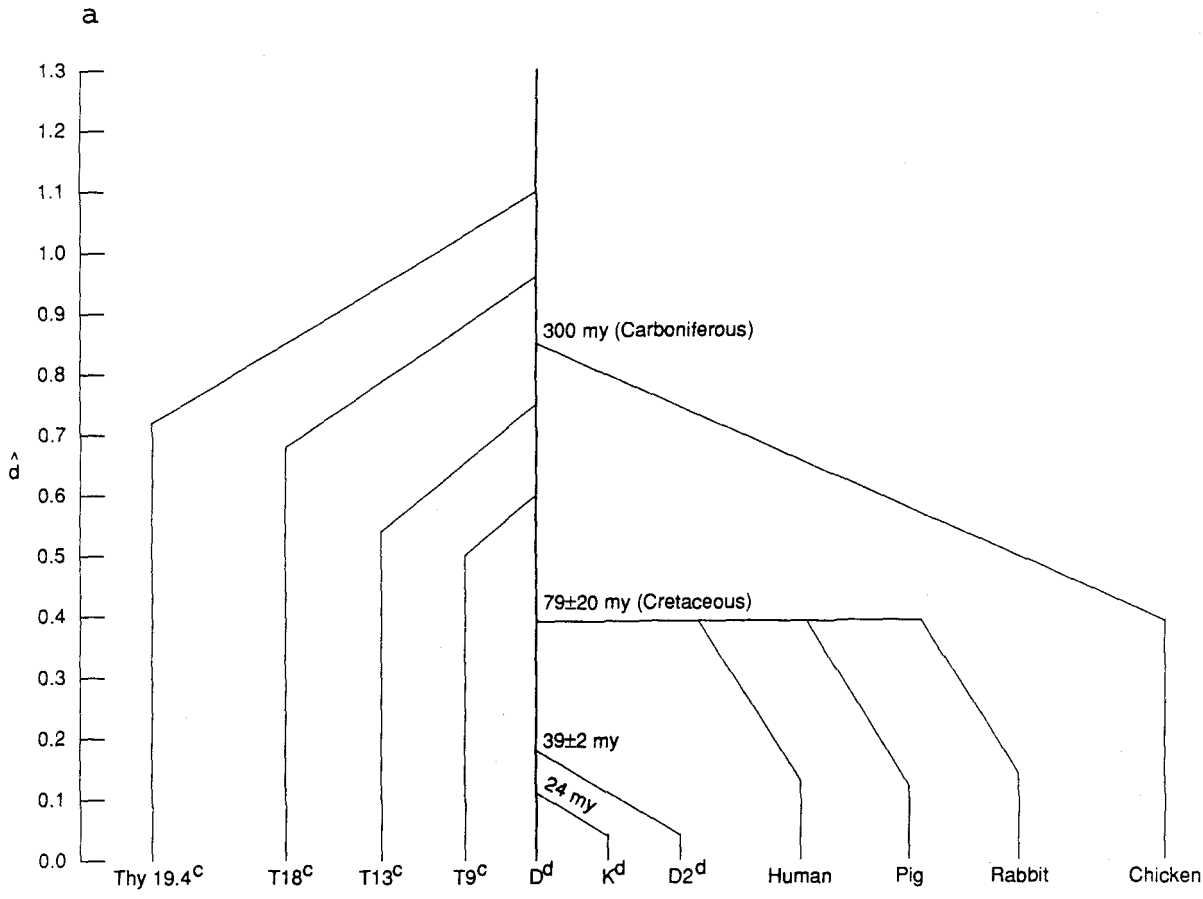
Genes Compared	Silent/Replacement Rates
Mammalian H4 Histones	357
Mammalian α-actins	262
Mammalian fibrinogens	10.6
Mammalian β_2-microglobulins	9.72
Mammalian α-fetoprotein	4.05
Mammalian β-globins	3.40
Mammalian α1 interferons	2.50
Exons 2 & 3 *Thy19.4* vs. D^d	4.5
Exons 2 & 3 $T13^c$ vs. D^d	2.3
Exons 2 & 3 $T18^c$ vs. D^d	2.1
Exons 2 & 3 $T9^c$ vs. D^d	1.4
Exons 2 & 3 *HLA-A2* vs D^d	1.4
Exons 2 & 3 *B-F* vs. D^d	1.4
Exons 2 & 3 K^d vs. D^d	0.92
Exon 4 *Thy19.4* vs. D^d	1.0
Exon 4 $T13^c$ vs. D^d	3.0
Exon 4 $T18^c$ vs. D^d	1.9
Exon 4 $T9^c$ vs. D^d	1.7
Exon 4 *HLA-A2* vs. D^d	2.2
Exon 4 *B-F* vs. D^d	1.4
Exon 4 K^d vs. D^d	0.86
Exon 4 *Ratmhc1* vs. D^d	2.0

Table adapted from Chapter 4, values for non-class I genes adapted from Li *et al.* (1985). *HLA-A2* is a human class I gene, while *B-F* is a chicken class I gene.

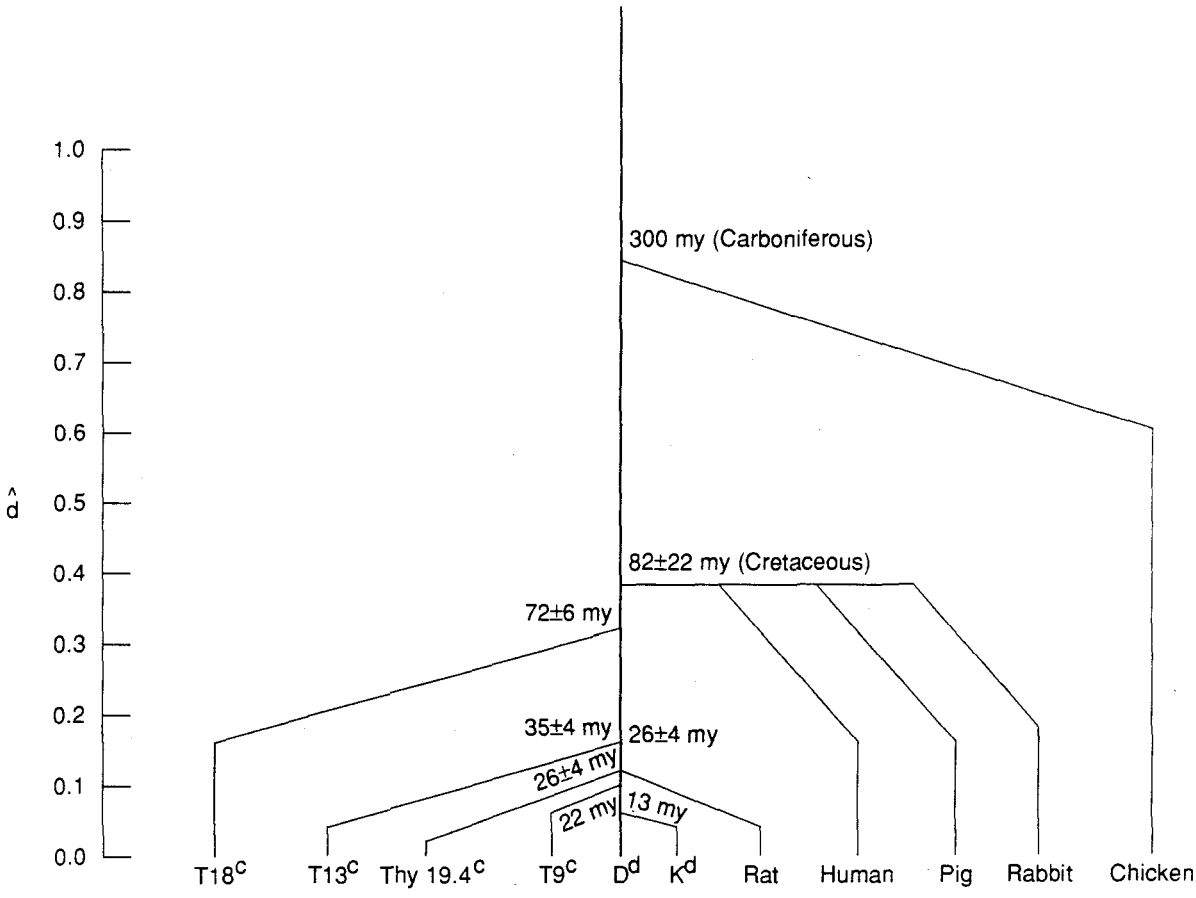
Fig. 1. *Phylogeny of Vertebrates and Other Organisms Based on Paleontological Evidence.* Branch points that date to the Precambrian era have not been resolved paleontologically. Mammals arose from synapsid reptiles in the Triassic period 200 million years ago (Lillegraven *et al.* 1979). Synapsid reptiles diverged from anapsid reptiles and their modern bird and reptile descendants in the Carboniferous period 300 million years ago (Young 1962). The divergence of the modern Eutherian orders occurred in the Cretaceous period 65-85 million years ago (Young 1962). Figure adapted from Nei *et al.* (1987).


PRE-CAM-BRIAN
CAMBRIAN
ORDO-VICIAN
SILU-RIAN
DEVO-NIAN
CARBONIF-EROUS
PER-MIAN
TRIAS-SIC
JURAS-SIC
CRETA-CEOUS
CENOZOIC
600
500
450
400
350
300
250
200
150
100
50
0 (Millions of years)
Therapsid Mammal like Reptiles
Earliest Mammals
Primates
Rodents
Rabbits
Whales and Porpoises
Carnivores (dogs, cats)
Artiodactyles (pigs, bovine)
Perissodactyls (horses)
Elephants
Marsupials
Dinosaurs
Birds
Crocodiles
Snakes
Lizards
Turtles
Frogs
Salamanders
Bony Fishes
Sharks and Rays
Lampreys and Hagfishes
Insects
Higher Plants
Fungi
Bacteria
?
?
?
?
?

Figs. 2a and b. ***Rooted phylogenetic trees of exon 2 and 3 as well as exon 4 evolution.*** Chicken, human, pig, rabbit, and rat sequences were included, since the divergence dates of mice from these vertebrates is known from the fossil record (Fig. 1). The second and third exons of the *Tla* region class I genes diverged from the transplantation antigens prior to the mammalian radiation, while their fourth exons appear to have diverged subsequently, probably because of gene conversion events. Figure adapted from Chapter 4.


a
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
d̂
300 my (Carboniferous)
79±20 my (Cretaceous)
39±2 my
24 my
Thy 19.4c
T18c
T13c
T9c
Dd
Kd
D2d
Human
Pig
Rabbit
Chicken

b


1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
d̂
300 my (Carboniferous)
82±22 my (Cretaceous)
72±6 my
35±4 my
26±4 my
26±4 my
22 my
13 my
T18c
T13c
Thy 19.4c
T9c
Dd
Kd
Rat
Human
Pig
Rabbit
Chicken

APPENDIX I

Transfectants of *Tla* Region Class I Genes

One of the major unresolved issues concerning the *Tla* region of the MHC is which class I genes contained within it encode proteins. The 19 *Tla* region class I genes were initially identified as sequences that hybridized to a fourth exon probe, pH2IIa. However, unlike the genes in the other regions of the MHC, the majority of the *Tla* region genes have not been characterized beyond that point. In fact, only one *Tla* region gene product, the TL molecule, has been studied extensively. To study protein expression of the *Tla* region gene family, an analysis was initiated to transfect each *Tla* region gene product into a fibroblast cell line. Constructs were created by subcloning each *Tla* region class I gene from the BALB/c class I cosmids into one of two expression vectors, pSV2gpt, or pUCSV2gpt (Table I; pUCSV2gpt kindly provided by B. Vernooy). These two vectors were chosen because they contain an SV40 enhancer element, which presumably would enhance the transcription of the transfected *Tla* region gene. In addition, a K^d construct was also created to serve as a positive control. To confirm that the complete and correct genes were subcloned into the individual constructs, the constructs were mapped with four restriction enzymes, BamHI, EcoRI, BglII, and HindIII as well as infrequent palindromic restriction enzymes originally used to map the class I MHC cosmids (Fig. 1). Exon four sequences were identified within the constructs by hybridization to the probe pH2IIa. Finally, in many cases subclones were derived from these constructs and sequenced in order to map individual exons. After the mapping analysis, only two of the twenty constructs, $T6^c$-pUCSV2gpt, and $T8^c$-pUCSV2gpt, were found not to contain complete class I genes.

These constructs were transfected, using the $CaPO_4$ method (Wigler *et al.* 1979), into SVCAS2.2F6, a tail fibroblast cell line from B10.CAS2 mice (cell line kindly provided by Dr. J. Rodgers, Baylor College of Medicine, Houston, TX). This

cell line was chosen because B10.CAS2 mice share the *H-2*^{w17}, *Qa-*^{w17}, *Tla*^{g}, and *Hmt*^{b} alleles with ***Mus musculus castaneous*** mice. Because of this, the probability of allelism between the transfected *Tla* region genes and the endogenous class I genes was presumed to be much higher than if a ***Mus musculus domesticus*** cell line is used. Thus, it was reasoned that protein products of the transfected genes could be distinguished more easily from endogenous molecules either by serological epitope differences or because of allelic molecular weight or charge differences. A fibroblast cell line was chosen because some *Tla* region genes are expressed in a lymphoid tissue-specific manner and thus fewer endogenous *Tla* gene products that could potentially interfere with detection of transfected gene products would be expected to be expressed in a fibroblast parental cell line than in a lymphoid parental cell-line. The transfectants were cotransfected with pSV2neo and selected for with G418 (Geneticin), since it was found that the transfectants were unstable in MXHAT (mycophenolic acid, xanthine, hypoxanthine, aminopterin, and thymidine) media. A killing curve revealed that 200 μg/ml of G418 are at least five times the amount required to kill the SVCAS2.2F6 parental cell-line, and thus that concentration was chosen for selection. Transfectants were generated using each *Tla* and *K* region class I gene construct (Table II).

Some of the transfectants were found to express their transfected gene at the protein or RNA level. A radioimmunoassay (RIA) using the 34-1-2 antibody (Ozato *et al.* 1982) demonstrated that the *K*^{d} gene transfectant, K^{d}A, has K^{d} molecules on its cell surface (Fig. 2). Fluorescent activated cell-sorting (FACS) analysis with TL.m4 antibodies (Shen *et al.* 1982) revealed that the *T13*^{c} gene transfectant, T13A, expresses TL^{C} determinants on its cell surface (Fig. 3). Finally, the *T9*^{c}, *T17*^{c}, and *T18*^{c} gene transfectants all appear to express their

transfected gene on the mRNA level (Fig. 4). However, the overall intent of identifying which *Tla* region class I genes could encode protein products did not prove feasible. This was partially a result of the limited available supply of serological reagents that are specific for *Tla* region gene products. Such reagents could be scarce possibly because the *Tla* region genes are not allelic between inbred mice, or because some of the genes are not expressed. To surmount this problem, an attempt was made to analyze radiolabeled whole-cell lysates of the transfectants for novel proteins using 2D protein gels run by a company, Protein Database, Inc. (Huntington Station, NY), which specializes in running gels with a high degree of uniformity. The K^d and $T13^c$ (TL^C) gene transfectants were chosen to serve as positive controls to check the feasibility of this approach. This approach unfortunately proved to be unreliable because of clonal variation of endogenous protein expression between the individual transfectants and because the transfected TL^C molecules could not be visualized in the positive control (T13A) cell lysate gels. This may have occurred because other endogenous proteins present at higher levels may have obscured them. Alternatively, differential glycosylation could make this molecule appear to be smear in the gel because of size or charge heterogeneity (J. Ketman, personal communication). Finally, an attempt to coprecipitate transfected class I molecules from radiolabeled lysates of the $T9^c$, $T17^c$, and $T18^c$ with antibodies specific for β_2-microglobin produced negative results, possibly because the putative molecules encoded by these genes and β_2-microglobulin associate with low avidity.

From this analysis, it appears that a more feasible approach to analyzing the expression of the *Tla* region genes is first to analyze the DNA sequence and mRNA expression of individual genes as described in Chapters 1 through 3. Once interesting genes, like $T18^c$ or *Thy19.4*, are identified, the transfectants can prove

useful for analyzing their possible gene products. Although the original intention of these transfectants did not prove feasible, the K^d and $T13^c$ gene transfectants have proven useful in other studies as positive controls (Cheroutre *et al.* in preparation; A. van Leeuwen, personal communication). It is hoped that the remainder will prove useful in future studies when more serological reagents become available for *Tla* region gene products, or if more extensive sequencing studies are initiated on the *Tla* region genes.

References

Fisher, D., Hunt, S., and Hood, L. 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**:528-545, 1985.

Fisher, D., Pecht, M., and Hood, L. DNA sequence of a class I pseudogene from the *Tla* region of the murine MHC: recombination at a B2 Alu repetitive sequence. *J. Mol. Evol.* **28**:306-312, 1989.

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

Ozato, K., Mayer, N., and Sachs, D. Monoclonal antibodies to mouse major histocompatibility complex antigens. *Transplantation* **34**:113-120, 1982.

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

Shen, F., Chorney, M., and Boyse, E. Further polymorphism of the *Tla* locus defined by monoclonal TL antibodies. *Immunogenetics* **15**:573-578, 1982.

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

Widmark, E., Ronne, H., Hamerling, U., Servenius, B., Larhammar, D., Gustafsson, K., Bohme, J., Peterson, P., and Rask, L. Family relationships of murine major histocompatibility genes: sequence of the $T2A^a$ pseudogene, a member of gene family 3. *J. Biol. Chem.* **263**:7055-7059, 1988.

Wigler, M., Pellicer, A., Silverstein, S., Axel, R., Urlaub, G. and Chasin, L. DNA-mediated transfer of the adenine phosphoribosyl transferase locus into mammalian cells. *Proc. Nat. Acad. Sci. USA* **76**:1373-1376, 1979.

Table I

Tla Region Gene Constructs

Gene	Source	Insert	Enzymes	Vector
$T1^c$	c66.1	12.2 kb	BamHI	pSV2gpt
$T2^c$	c59.1	5.4 kb	XbaI	pUCSV2gpt
$T3^c$	c52.2	14.6 kb	BamHI	pSV2gpt
$T4^c$	c9.1	10.6 kb	BamHI	pSV2gpt
$T5^c$	c9.1	8.8 kb	SmaI	pUCSV2gpt
$T6^c$	c5.2	6.8 kb	SmaI	pUCSV2gpt
$T7^c$	c5.2	9.0 kb	SmaI	pUCSV2gpt
$T8^c$	c48.1	8.8 kb	NruI	pUCSV2gpt
$T9^c$	c48.1	8.6 kb	HpaI	pUCSV2gpt
$T10^c$	c12.2	10.4 kb	HpaI	pUCSV2gpt
$T11^c$	c22.1	12.4 kb	BamHI	pSV2gpt
$T12^c$	c22.1	5.4 kb	XbaI	pUCSV2gpt
$T13^c$	pTLA.1	12.0 kb	ClaI/HpaI	pUCSV2gpt
$T14^c$	c6.3	9.6 kb	HpaI	pUCSV2gpt
$T15^c$	c6.3	7.4 kb	SmaI	pUCSV2gpt
$T16^c$	c6.3	11.0 kb	HpaI/XhoI	pUCSV2gpt
$T17^c$	c37-16	7.4 kb	EcoRI	pUCSV2gpt
$T18^c$	c15.3	11.4 kb	BamHI	pSV2gpt
37^c	c37-16	8.0 kb	EcoRI/SalI	pUCSV2gpt
K^d	λ1.3	11.0 kb	EcoRI	pUCSV2gpt
5' #1 (weak)	c66.1	10.6 kb	BamHI	pSV2gpt

Table II

Tla Region Transfectants

Gene	Gene Product	Transfectants	Status of Transfectants
T1^c	ψ gene	T1.2, T1.1	
T2^c	ψ gene	T2D	
T3^c		T3.2, T3.4, T3.5	
T4^c		T4A, T4B, T4C, T4D	
T5^c	ψ gene	T5A, T5C	
T6^c		-	
T7^c		T7A, T7B, T7.5	
T8^c		-	
T9^c	ψ gene	T9C, T9.13, T9.14 T9.15, T9.16	All express T9 mRNA
T10^c	ψ gene	T10B, T10D	
T11^c	ψ gene or secreted	T11C, T11D, T11.5	
T12^c	ψ gene or secreted	T12C, T12B	
T13^c	TL^c molecule	T13A, T13B	T13A expresses TL^c determinants on cell surface
T14^c		T14A, T14B, T14.2, T14.3	
T15^c		T15A, T15C	
T16^c		T16.1, T16.4, T16.6 T16A, T16B	
T17^c		T17.3, T17.4, T17.5 T17.6	All express T17 mRNA
T18^c		T18.1, T18.2, T18.5	All express T18 mRNA
		T18C, T18D	
37^c		37.3, 37.4, 37.5, 37.6	
K^d	H-2K^d	K^{d}A	K^{d}A expresses K^d molecules on cell surface

Figure 1. *Restriction maps of 20 class I gene constructs.* Restriction endonuclease abbreviations are B for BamHI, G for BglII, R for EcoRI, and H for HindIII. Fourth exon sequences are mapped within these constructs by hybridization to the probe pH2IIa (Steinmetz *et al.* 1982), while other exons are mapped by sequence of subclones indicated by number above the construct map. Genes that were mapped or sequenced before this study are: K^d (Kvist *et al.* 1983); $T1^c$ (Fisher *et al.* 1989); *T2* (Widmark *et al.* 1988); $T5^c$ and $T8^c$ (Rogers 1985); $T13^c$ (Fisher *et al.* 1985); $T9^c$, $T17^c$ and $T18^c$ (Chapter 3). Dashed lines connect presumed homologous restriction endonuclease sites between duplicate genes.

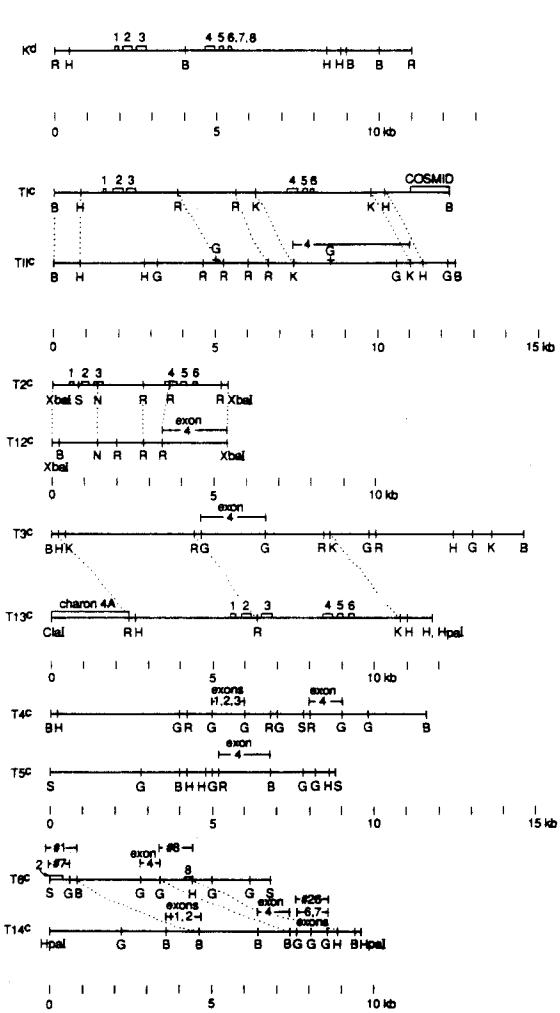

Kd
1 2 3
4 5 6,7,8
R H
B
H H B
B
R
0
5
10 kb
T1c
1 2 3
4 5 6
COSMID
B H
R
R K
K H
B
G
4
G
T11c
B H
H G
R R R R K
G K H G B
0
5
10
15 kb
T2c
1 2 3
4 5 6
Xbal S N
R R
R Xbal
exon
4
T12c
B
N R R R
Xbal
Xbal
0
5
10 kb
exon
4
T3c
B H K
R G
G
R K
G R
H G K B
charon 4A
1 2 3
4 5 6
T13c
ClaI
R H
R
K H H, HpaI
0
5
10 kb
exons
H.2,3
exon
4
T4c
B H
G R G G R G S R G G B
exon
4
T5c
S
G
B H H G R
B
G G H S
0
5
10
15 kb
#1
2
#7
H
exon
4
#8
8
T6c
S G B
G G
H G
G S
exons
H.1,2
exon
4
#26
6,7
exons
T14c
HpaI
G
B B
B
B G G G H B HpaI
0
5
10 kb
#20-17
#13
T7c
R
R
5 6
S
G G
G G G G G G
B
B S
exons
#34
#36
6,7
1,2,3
4
5
T15c
B R H,R G
G G G G H
B
#34
0
5
10 kb
exon
3
exon
4
T8c
N R B H
G
B B R
exon
4
H R G N
1,2
3
exons
exon
R
R
T16c
XhoI S H
H B
S
R
B H
N
0
5
10
15 kb
T9
2
4 5
HpaI
B
K
B
HpaI
T17
2
4 5
R B
K
B
R
0
5
10 kb
exon
4
T10
HpaI R B H
H
G H B
SacII
B HpaI
1 2 3
4 5 6 7
vector
37
R B
B
B
SalI
0
5
10 kb
exon
4
T18
cosmid
exon 3
B
K
S K
HpaI
B
portion
5'
5'#1
B G H G
S R H
R
S
G B
0
5
10
15 kb

Figure 2. *Radioimmunoassay using H-2K^d specific monoclonal antibody 34-1-2.* At all titrations of the antibody, the K^d gene transfectant cells have at least five times as many radioactive antibodies on their surface as the parental SVCAS2.2F6 cells.

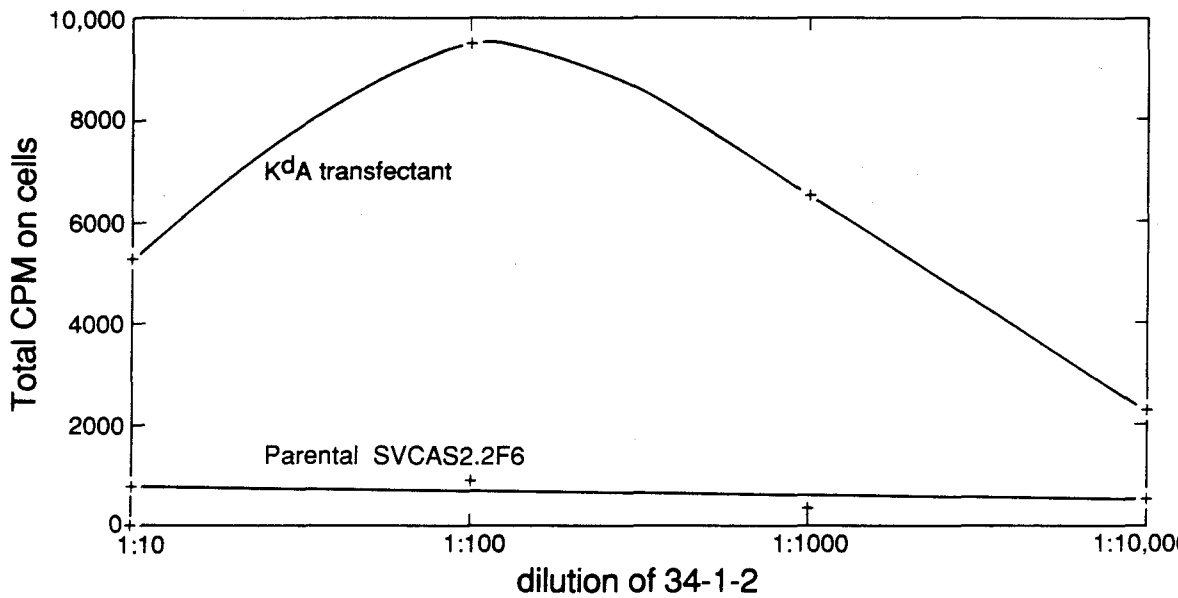
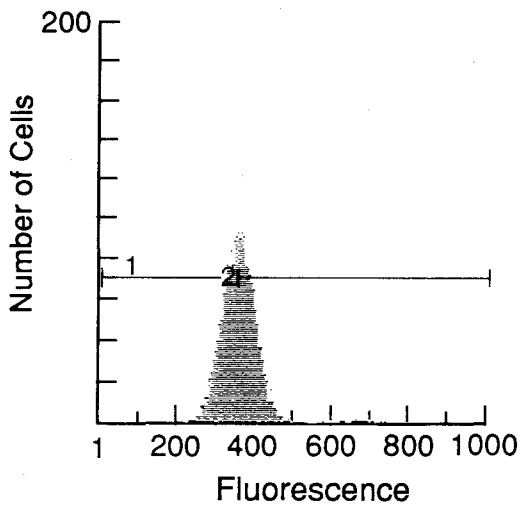
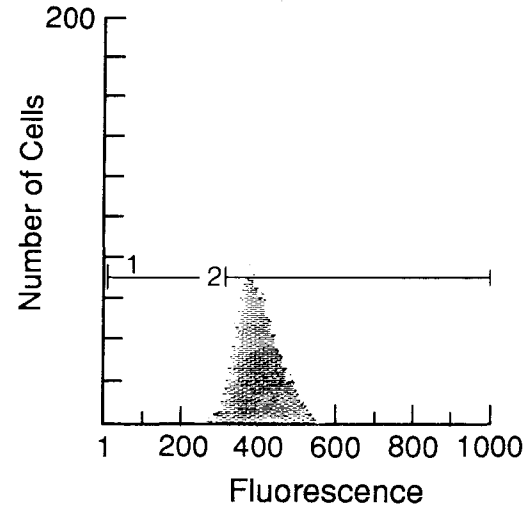

10,000
8000
6000
4000
2000
0
Total CPM on cells
KdA transfectant
Parental SVCAS2.2F6
1:10
1:100
1:1000
1:10,000
dilution of 34-1-2

Figure 3. *FACS analysis using TL^c specific monoclonal antibody TL.m4.* The *$T13^c$* gene transfectant T13A has a significant peak shift relative to the parental SVCAS2.2F6 cell line.



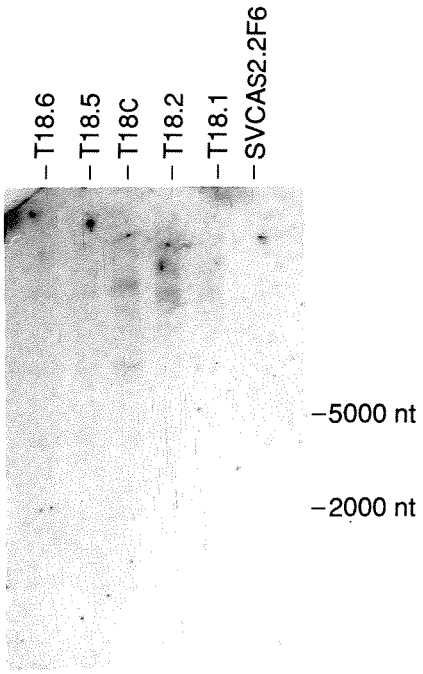
SVCAS2.2F6
Parental



T13A
Transfectant

Figure 4. *Northern blot analysis of $T9^C$, $T17^C$ and $T18^C$ transfectants.* Northern blots were probed with $T9^C/T17^C$ and $T18^C$ gene-specific cDNA probes (Chapter 3). Visible bands are evident in almost all transfectants, although more apparent in $T9^C$ and $T17^C$ gene transfectants.

Probed with T18 5´ probe



Probed with T9/T17 3´ probe