CLASS I GENES OF THE MAJOR HISTOCOMPATIBILITY COMPLEX: STRUCTURAL STUDIES ON GENES OF THE *Tla* LOCUS

Thesis by

Douglas Arthur Fisher

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1986

(Submitted September 19, 1985)

To my mother

for her love and encouragement.

Acknowledgements

I would like to thank my advisor, Lee Hood, for the training I received in his laboratory. He is an outstanding scientist with seemingly boundless enthusiasm.

My thanks to the many professors, postdocs, and graduate students with whom I have argued, listened, and learned. I am grateful to my collaborators, Michael Steinmetz, John Frelinger, Marit Pecht, and Steve Hunt for their contributions to this thesis, to Karyl Minard and Debbie Maloney for pouring the countless sequencing gels, and to Bertha Jones, Ria DeBruyn, and Jessie Walker for keeping the lab running smoothly. I enjoyed the interesting noontime discussions at "The Greasy" with Lloyd Smith, Paul Tempst, Lance Fors, and Jerry Siu on such wide-ranging topics as softball, politics, and the feasibility of building a time machine.

I am very grateful to Connie Katz and the secretarial staff for typing this thesis quickly, accurately, and with a smile.

I would like to thank the National Institutes of Health for financial support and the Weigle Fund for support in preparing this thesis.

Thanks to my good friend David Bassham, who likes to see his name in print.

Lastly, my very special thanks to my mother for her faith in me, her love, and her butterscotch brownies, and to my fianceé, Mary, whose friendship and understanding has helped me get through and whose love is an inspiration.

Abstract

This thesis contains investigations into the structure of molecules encoded within the mouse major histocompatibility complex. The first chapter [Steinmetz, M., J. G. Frelinger, D. Fisher, T. Hunkapiller, D. Pereira, S. M. Weissman, S. G. Nathenson, and L. Hood, *Cell* **24:** 125] describes the isolation and characterization of the first cDNA clones encoding murine transplantation (H-2) antigens. This study showed that H-2 antigens contain DNA and protein sequences related to immunoglobulin (Ig) molecules, but that the similarity does not include the great somatic diversity characteristic of Ig molecules.

The second chapter contains methods for cloning and sequencing in M13 bacteriophage vectors. Included is a novel method of generating overlapping subclones for DNA sequencing by making a family of deletions in a DNA insert cloned in M13.

In the third chapter [Fisher, D. A., S. W. Hunt, and L. Hood *J. Exp. Med.* **162:** 528], the complete structure of a gene encoding a serologically defined thymus leukemia (TL) antigen is elucidated. TL antigen is encoded in a gene, gene $T13^C$, closely related to H-2 antigens, and appears to have undergone a gene conversion event with an H-2 gene. *Tla*-specific probes subcloned from $T13^C$ enabled us to examine the organization of the eighteen cross hybridizing class I genes of the *Tla* region.

The last chapter contains the sequence of another gene, gene $T1^C$, previously identified as encoding TL antigen. However, the $T1^C$ gene is a non-functional pseudogene, and was probably mis-identified. There is an apparent site of recombination in the $T1^C$ gene that occurs precisely at a B2 Alu repeat sequence.

Table of Contents

Introduction

The murine class I antigens are a family of structurally related cell surface glycoproteins encoded within the major histocompatibility complex (MHC) of chromosome 17 (1, 2). These antigens include the transplantation antigens from the *H-2* complex, H-2K, H-2D, and H-2L, as well as the lymphoid differentiation antigens Qa-1, Qa-2, 3, and thymus leukemia (TL) antigen encoded in the *Qa-1, Qa-2, 3,* and *Tla* loci (3-6). Transplantation antigens, first discovered because of their ability to provoke graft rejection in allogeneic skin grafts (7, 8), actually function as cell-cell recognition molecules in the killing of virus-infected cells by cytotoxic T lymphocytes (9). The Qa/TL antigens do not elicit a strong cellular immune response and have no known function, but their presence on lymphoid cells, coupled with their homology to H-2 antigens, suggests that they too may be involved in cell-cell interactions in the immune system.

Class I molecules: genetics and structure. The MHC encodes at least three classes of immune system related molecules (Fig. 1). The I-region (class II) antigens, expressed only on lymphoid cells, are involved in communications between lymphocytes that regulate immune responsiveness (10). Class III molecules, encoded within the S-region, are components of the complement pathway (11). Class I molecules share structural characteristics, but may differ in their polymorphism and patterns of expression: H-2 antigens are expressed on virtually all somatic cells and are extraordinarily polymorphic, with over 100 serologically defined alleles reported at both the *H-2K* and *H-2D* loci (12). In contrast, only two alleles of the *Qa-1* and *Qa-2, 3* loci exist (one of which encodes the absence of antigen), and these molecules are expressed only on lymphoid cells (3).

Class I molecules are ~45,000 dalton transmembrane glycoproteins that associate non-covalently with a 12,000 dalton polypeptide, β_2-microglobulin. The

general structure of a class I molecule is shown in Fig. 2. There are three external protein domains of about 90 amino acids each, denoted α1, α2, and α3. A hydrophobic segment spans the lipid bilayer, and a small hydrophilic domain is located on the cytoplasmic side of the membrane (13).

The membrane proximal external domain, α3, has a highly conserved protein sequence among all class I molecules, and is the binding site for β_2-microglobulin (14). The protein sequence of the α3 domain has significant homology to that of immunoglobulin (Ig) domains (15, 16), a fact that probably reflects its molecular complementarity with β_2-microglobulin, since β_2-microglobulin is itself related to Ig domains and has been termed a "free immunoglobulin domain" (17). Although H-2 antigens show extensive allelic polymorphism, they do not share the enormous somatic diversity associated with the variable regions of immunoglobulin molecules (16).

The α1 and α2 domains are the most polymorphic part of the molecule when alleles of H-2 antigens are compared. It is not surprising that the antigenic determinants recognized by antibodies and cytotoxic lymphocytes are virtually all in the α1 and α2 domains (18-21). Recent X-ray crystallographic studies suggest that the α1 and α2 domains fold to form a larger "domain," as do the α3 and β_2-microglobulin domains (22). Thus a class I molecule consists of a polymorphic N-terminal region and a conserved membrane proximal region composed of two polypeptide chains.

Thymus leukemia antigen and the Tla locus. Thymus leukemia (TL) antigen was originally discovered by immunizing C57BL mice with radiation-induced and spontaneous leukemias (4). The resulting antisera defined an antigen present on some C57BL leukemias but on no normal C57BL tissue. Leukemias of other mouse strains, like the A strain, also reacted with the antibody; in addition, A strain thymocytes (but no other tissue) bound the antibody. For this reason, the A strain

was designated as TL^+ and C57BL as TL^-. The finding that TL^- mice could have TL^+ leukemias led Boyse and Old (23) to suggest that TL^- mice have both a TL structural gene and a regulatory gene that represses TL expression. In the leukemic state the repression would be removed and TL antigen expressed. The role of TL antigen in the leukemic state is unknown.

The locus encoding TL antigen, *Tla*, was mapped to within 2 centiMorgans of the *H-2* locus (Fig. 1) (24). Currently there are six serologically defined alleles of the *Tla* locus, designated Tla^{a-f} (25, 26). The proximity of the *Tla* locus to the *H-2* locus is not coincidental, as TL antigens are structurally similar to H-2 antigens. TL antigens have a molecular weight of ~45,000 daltons and associate with β_2-microglobulin (2, 3). Peptide map comparisons (27, 28) confirmed the relatedness of H-2 and TL antigens; however, H-2 antigens are more closely related to each other than to TL antigens.

The advent of recombinant DNA technology has enabled purified genes to be examined in great detail, and to be manipulated experimentally. Work described in this thesis includes the cloning of murine class I sequences (16). These clones were used as probes to isolate numerous genomic class I genes (12). Functional genes, including one encoding TL antigen, were located by DNA mediated gene transfer of the clones into mouse L cells (29). The remainder of this thesis concerns the structural analysis of two class I genes from the *Tla* locus of the BALB/c mouse, one of which encodes a serologically defined TL antigen.

The rationale behind these studies is to better understand molecules significant in the functioning of the immune system. We know that H-2 molecules are functionally significant, but what is the role of the structurally related antigen, TL, expressed specifically in the thymus, the very organ in which H-2 restriction is learned? Structural information on *Tla* genes will enable us to access their relatedness to *H-2* genes, and will answer questions regarding the

evolution of the class I gene family. Gene specific probes isolated from the clones *Tla* gene should allow one to study the expression of *Tla* sequences in RNA, and will allow other *Tla* alleles with different patterns of expression to be cloned. Finally, having purified *Tla* genes permits one to manipulate those genes by introducing them in biologically unfamiliar locations or to construct hybrid genes to map antigenic or functional parts of the molecule.

References

1. Klein J. 1975. Biology of the mouse histocompatibility-2 complex. (Berlin: Springer-Verlag).
2. Vitetta, E. S., and J. D. Capra. 1978. The protein products of the murine 17th chromosome: genetics and structure. *Adv. Immun.* **26**:147.
3. Flaherty L. 1980. The Tla-region antigens. *In* the Role of the Major Histocompatibility Complex in Immunology. M. E. Dorf, ed. Garland Press, New York, pp. 33-57.
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. Gorer, P. A. 1938. The antigenic basis of tumour transplantation. *J. Pathol. Bacteriol.* **47**:231.
8. Gorer, P. A., S. Lyman, and G. Snell. 1948. Studies on the genetic and antigenic basis of tumor transplantation: Linkage between a histo-compatibility gene and "fused" in mice. *Proc. R. Soc. (Lond) B* **135**:499
9. 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.
10. Kaufman, J. F., C. Auffray, A. J. Korman, D. A. Shackelford, and J. Strominger. 1984. The class II molecules of the human and murine major histocompatibility complex. *Cell* **36**:1.

11. Schreffler, D. C. 1976. The S region of the mouse major histocompatibility complex (H-2): genetic variation and functional role in complement system. *Transplantation Rev.* **32**:140.

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

13. Nathenson, S. G., H. Uehara, B. M. Ewenstein, T. J. Kindt, and J. E. Coligan. 1981. Primary structural analysis of the transplantation antigens of the murine H-2 major histocompatibility complex. *Ann. Rev. Biochem.* **50**:1025.

14. Yokoyama, K., and S. G. Nathenson. 1983. Intramolecular localization of class I H-2 MHC antigens; localization of the allogenic determinants and β_2-m binding site to different regions of the H-2k^b glycoprotein, *J. Immunol.* **130**:149.

15. Orr, H. T., D. Lancet, R. Robb, J. A. Lopez de Castro, and J. L. Strominger. 1979. The heavy chain of human histocompatibility antigen HLA-B7 contains an immunoglobin-like region *Nature* **232**:266.

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

17. Peterson, P. A., B. A. Cunningham, I. Berggard, and G. M. Edelman 1972. β_2-microglobulin -- a free immunoglobulin domain. *Proc. Natl. Acad. Sci. USA* **69**:1697.

18. Evans, G. A., D. H. Margulies, B. Shykind, J. G. Seidman, and K. Ozato. 1982. Exon shuffling: mapping polymorphic determinants on hybrid mouse transplantation antigens. *Nature* 300:755.

19. Murre C., E. Choi, J. Weis, J. G. Seidman, K. Ozato, L. Liu, S. J. Burakoff, and C. S. Reiss. 1984. Dissection of serological and cytolytic T lymphocyte

epitopes on murine major histocompatibility antigens by a recombinant H-2 gene separating the first two external domains. *J. Exp. Med.* **160:**167.

20. Reiss C., G. A. Evans, D. H. Margulies, J. G. Seidman, S. J. Burakoff. 1983. Allospecific and virus-specific cytolytic T lymphocytes are restricted to N or Cl domain of H-2 antigens expressed on L cells after DNA-mediated gene transfer. *Proc. Natl. Acad. Sci. USA* **80:**2709.
21. Stroynowski, I., A. Orn, R. S. Goodenow, M. McMillan, J. Forman, P. R. Brayton, J. Frelinger, and L. Hood. 1984. Cytotoxic T lymphocytes recognize determinants on the BALB/c H-2L^{d} molecule controlled by α1 and α2 but not α3 external domains. *Immunogenetics* **20:**141.
22. Björkman, P. 1984. Ph.D. Thesis, Harvard University.
23. Boyse, E. A. and L. L. Old. (1969). Some aspects of normal and abnormal cell surface genetics. *Ann. Rev. Genetics* **3:**269.
24. Boyse, E. A., L. J. Old, and S. Luell. (1964). Genetic determination of the TL (thymus-leukemia) antigen in the mouse. *Nature* **201:**779.
25. Shen, F.-W., M. Chorney, and E. A. Boyse. (1982). Further polymorphism of the Tla locus defined by monoclonal antibodies. *Immunogenetics* **15:**573.
26. Michaelson, J., E. A. Boyse, M. Chorney, L. Flaherty, E. Fleissner, U. Hämmerling, C. Reinisch, R. Rosenson and F.-W. Shen. (1983). The biochemical genetics of the Qa-Tla region. *Transplantation Proc.* **15:**2033.
27. Yokoyama, K., E. Stockert, L. J. Old, and S. G. Nathenson. 1981. Structural comparisons of TL antigens derived from normal and leukemia cells of TL^{+} and TL^{-} strains and relationship to genetically linked H-2 major histocompatibility complex products. *Proc. Natl. Acad. Sci. USA* **78:**7078.
28. McIntyre, K. R., U. Hämmerling, J. W. Uhr, and E. S. Vitetta. 1982. Structural analysis of the thymus leukemia (TL) antigens in the mouse *J. Immunol.* **128:**1712.

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

Figure 1. Genetic map of the MHC. The order of loci within brackets is not known.

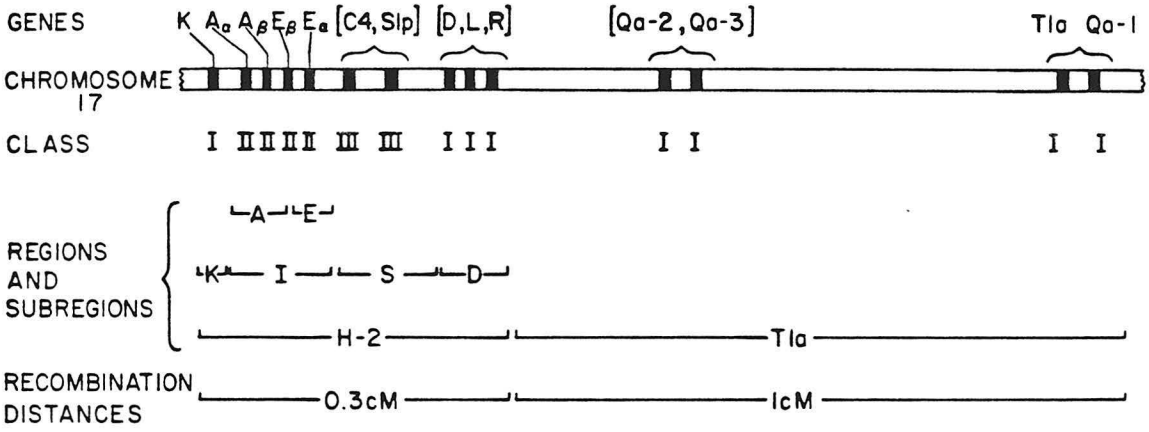
GENES
K Aα Aβ Eβ Eα [C4, Slp] [D,L,R] [Qa-2, Qa-3] Tla Qa-1
CHROMOSOME 17
CLASS
I IIIIIII III III III I I I I I
REGIONS AND SUBREGIONS
A E
K I S D
H-2
Tla
RECOMBINATION DISTANCES
0.3cM
1cM

Figure 1

Figure 2. Structure of a class 1 molecule.

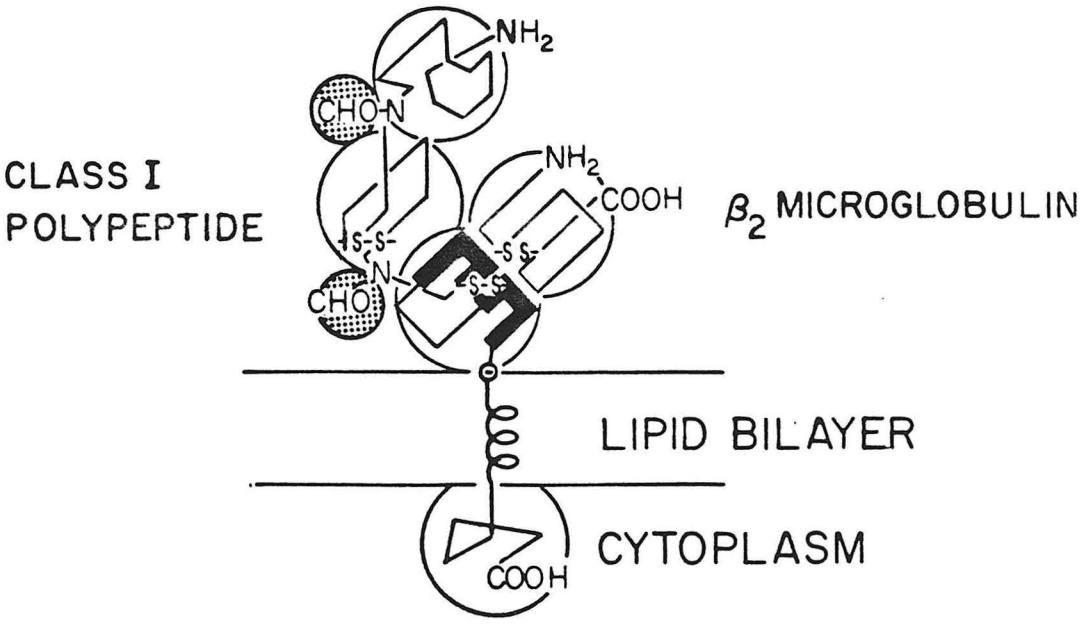
CLASS I
POLYPEPTIDE
β_2 MICROGLOBULIN
NH_2
CHO-N
NH_2
COOH
S-S
S-S
S-S
N
CHO
LIPID BILAYER
CYTOPLASM
COOH

Figure 2

Chapter 1

THREE cDNA CLONES ENCODING MOUSE TRANSPLANTATION ANTIGENS: HOMOLOGY TO IMMUNOGLOBULIN GENES

By Michael Steinmetz, John G. Frelinger, Douglas Fisher, Tim Hunkapiller, Dennis Pereira*, Sherman M. Weissman*, Hiroshi Uehara+, Stanley Nathenson+, and Leroy H. Hood

Division of Biology
California Institute of Technology
Pasadena, California 91125
*Department of Human Genetics
Yale University School of Medicine
New Haven, Connecticut 06510
+Departments of Microbiology and Immunology and Cell Biology
Albert Einstein College of Medicine
Bronx, New York 10461

Running title: H-2 cDNA Clones

Cell **24**, pp. 125-134

Michael Steinmetz,* John G. Frelinger,* Douglas Fisher,* Tim Hunkapiller,* Dennis Pereira,† Sherman M. Weissman,† Hiroshi Uehara,‡ Stanley Nathenson,‡ and Leroy Hood*

* Division of Biology
California Institute of Technology
Pasadena, California 91125
† Department of Human Genetics
Yale University School of Medicine
New Haven, Connecticut 06510
‡ Departments of Microbiology and Immunology and Cell Biology
Albert Einstein College of Medicine
Bronx, New York 10461

Summary

We constructed cDNA libraries from poly(A)$^+$ RNA isolated from cell lines of two different inbred strains of mice, and screened the libraries with a cDNA clone encoding a human transplantation antigen. Three cDNA clones were identified, sequenced and found to encode amino acid sequences highly homologous to portions of a known mouse transplantation antigen. Comparison of the cDNA sequences of mouse transplantation antigens with the constant region domains of the mouse immunoglobulin μ gene reveals a striking homology, which suggests that the two genes share a common ancestor. Antibody genes undergo DNA rearrangements during B cell differentiation that are correlated with their expression. In contrast, DNA blots with these cDNA probes suggest that the genes for the transplantation antigens are not rearranged in the genomes of liver or embryo cells, which express these antigens, as compared with sperm cells, which do not express these antigens. In Bam HI–digested liver DNAs from different inbred strains of mice, 10–15 bands of hybridization were found. Accordingly, the genes encoding the transplantation antigens appear to constitute a multigene family with similar gene numbers in different mice.

Introduction

The major histocompatibility complex of mammals is a tightly linked cluster of genes encoding a variety of proteins involved in the immune response. One family of these molecules is denoted the transplantation or histocompatibility antigens because differences in these proteins can cause rapid graft rejection (Gorer, 1938; Gorer et al., 1948). Transplantation antigens are found on the cell surfaces of all mammalian somatic cells, albeit at varying concentrations. They play a fundamental role in T cell surveillance mechanisms, which recognize virally infected or in some cases neoplastically transformed cells. Indeed, cytotoxic T cells are restricted to killing cells that express both a foreign determinant, such as a viral antigen, and a self-transplantation antigen (for review see Shearer and Schmitt-Verhulst, 1977; Zinkernagel and Doherty, 1980).

The mouse is an ideal subject for the study of genes encoding transplantation antigens because detailed serological and genetic analysis of the mouse major histocompatibility (H-2) complex has been facilitated by the existence of inbred, congenic and recombinant strains (Klein, 1975; Snell et al., 1976). Some mice appear to have at least four major transplantation antigens, K, D, L and R—all encoded on chromosome 17 (Hansen et al., 1981). The K and D genes are approximately 0.5 centimorgans apart while L and R are closely linked to D, but their relative positions have not yet been determined. Individual alleles of the H-2 complex are linked together in a large number of distinct combinations called haplotypes. The haplotype of an inbred mouse strain is denoted by a small letter, for example, d for the inbred BALB/c strain. The transplantation antigens of BALB/c mice are denoted as the K^d, D^d, L^d and R^d molecules. Transplantation antigens are extremely polymorphic by serological analyses; for example, in the mouse there are at least 56 alleles at the K locus and 45 alleles at the D locus (Klein, 1979).

The transplantation antigens of mice and other species exhibit homologous structures comprising two polypeptide chains—an integral membrane glycoprotein of approximately 45,000 daltons that is noncovalently associated with a 12,000 dalton component, β_2–microglobulin (Vitetta and Capra, 1978). Hereafter we shall use the term transplantation antigen to refer only to the 45,000 dalton component. The gene for the 45,000 dalton component is encoded in the H-2 complex, whereas the gene for the mouse β_2–microglobulin has not yet been mapped.

The complete amino acid sequence of 346 residues for the K^b transplantation antigen has been determined (Martinko et al., 1980; Uehara et al., 1981). Partial amino acid sequences of other mouse K, D and L transplantation antigens demonstrate that they are approximately 80% homologous to one another (Nathenson et al., 1981). The extensive serological polymorphism among molecules encoded by different alleles is reflected in extensive amino acid substitutions (Maizels et al., 1978; Coligan et al., 1980; Nairn et al., 1980; Nathenson et al., 1981). In addition, amino acid sequence comparisons suggest that portions of the transplantation antigens may be homologous to immunoglobulin molecules (Orr et al., 1979; Strominger et al., 1980).

Two groups of investigators have cloned cDNA probes for human transplantation antigens (Ploegh et al., 1980; Sood et al., 1981). We report the characterization of three cDNA clones encoding three distinct

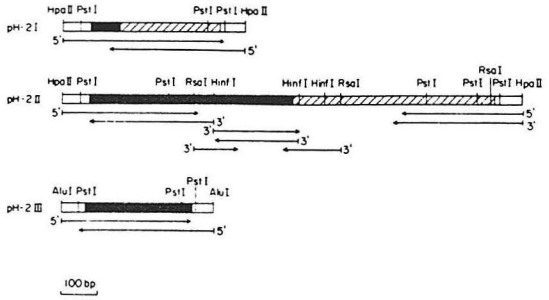

Figure 1. Partial Restriction Maps for the cDNA Clones pH-2I, pH-2II and pH-2III

Restriction sites were mapped on the cDNA inserts of the three plasmid DNAs by single and double restriction enzyme digestions in order to facilitate DNA sequence analysis. Arrows: extent and direction of DNA sequencing; 5′ or 3′: labeled ends. Open boxes: pBR322 sequences and G–C tails; filled boxes: coding sequences; hatched boxes: 3′ untranslated regions.

mouse transplantation antigens and demonstrate a significant homology relationship between a portion of the genes encoding the transplantation antigens and immunoglobulin constant region domains.

Results and Discussion

Three Mouse cDNA Clones Encode Molecules Which Are Highly Homologous to the K^b Transplantation Antigen

Using a human HLA cDNA clone (Sood et al., 1981) as a probe, we screened two cDNA libraries that were constructed from poly(A)$^+$ RNA from two mouse lymphoma cell lines. The C14 cell line, induced in a BALB/c (d haplotype) mouse by Abelson virus, overproduces the D^d molecule (Nairn et al., 1980), and the lymphoma cell line RDM-4 (k haplotype) overproduces the K^k antigen (Herrmann and Mescher, 1979). Two

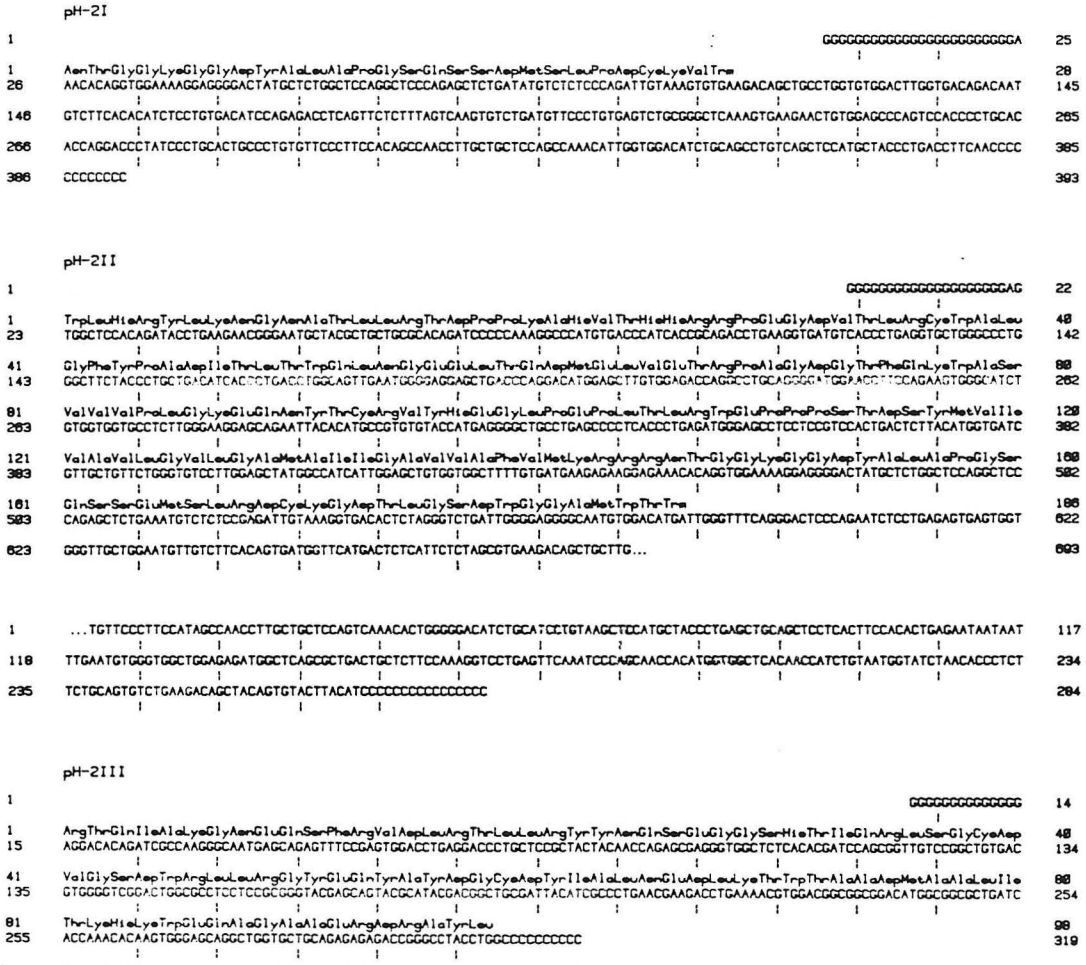

Figure 2. DNA Sequences for the Inserts of the cDNA Clones pH-2I, pH-2II and pH-2III

The complete DNA sequences of the noncoding strands for the inserts of clones pH-2I and pH-2III are given together with the predicted amino acid sequences encoded by the cDNA clones. For clone pH-2II the sequence is given except for a stretch of about 170 bp in the 3′ untranslated region. When compared with the completely sequenced K^b transplantation antigen (see Figure 3), clone pH-2I encodes amino acids 313 to 339, clone pH-2II codes for amino acids 167 to 352, and clone pH-2III encodes amino acids 63 to 160.

cDNA clones (pH-2I and pH-2II) were isolated from the C14 library and one cDNA clone (pH-2III) was isolated from the RDM-4 library. The approximate sizes of the inserts ranged between 300 and 1150 nucleotides: pH-2I, ~400 bp; pH-2II, ~1150 bp; and pH-2III, ~300 bp.

In order to characterize these clones, we determined the nucleotide sequences of the inserts for the three cDNA clones. Figure 1 gives partial restriction maps of the three cDNA clones and shows the DNA sequencing strategy employed in each case. Figure 2 gives DNA sequences obtained for these clones. These three DNA sequences were translated into amino acid sequences in all possible reading frames and then were compared with the fully determined amino acid sequence of the K^b molecule (Figure 3). Unambiguous homologies between the K^b molecule and the amino acid sequences were found in one reading frame for each of the three cDNA clones. Amino acid sequences in the correct reading frame are given in Figure 2.

The three isolated cDNA clones encode different portions of the transplantation antigen (Figure 3). Clone pH-2III codes for amino acids 63–160 and is 89% homologous to the K^b sequence. Clone pH-2II codes for amino acids 167–352, is 80% homologous to the K^b molecule and contains about 600 nucleotides of the 3′ untranslated region. Clone pH-2I codes for 27 amino acids at the C terminus, is 89% homologous to the K^b molecule and extends 274 nucleotides into the 3′ untranslated region.

We have compared the translated sequences for these three cDNA clones against the available protein sequence data for transplantation antigens derived from mice of the k and d haplotypes (Nairn et al., 1980; Rothbard et al., 1980; J. E. Coligan, personal communication). These comparisons allow us to exclude certain possibilities (Table 1) and to conclude that each of these cDNA clones encodes a distinct transplantation antigen, but no unambiguous assignments can be made, in part because of the paucity of amino acid sequence data available for the transplantation antigens (for example, no sequence data are available on the R^d molecule). Accordingly, it will be important to obtain protein sequence data as well as DNA data for subsequent gene and protein correlations in this system.

The pH-2II cDNA Clone Contains Repetitive DNA Sequences

When we used the pH-2II cDNA clone to analyze mouse DNA cleaved with Eco RI by Southern blot hybridization, this probe hybridized to a large number of genomic DNA fragments (Figure 4A). This type of hybridization indicates that the corresponding cDNA clone contains one or more repetitive elements (Steinmetz et al., 1980). To localize the repetitive element(s) on the cDNA clone, total mouse DNA was used as a probe against pH-2II DNA cleaved with various restriction enzymes, under conditions where only repetitive sequences will hybridize (Figure 4B). The repetitive sequence was mapped to the 3′ end of the pH-2II cDNA sequence. Indeed, restriction fragments from mouse DNA hybridized with two different intensities,

Table 1. cDNA Clones for Transplantation Antigens

Cell Line	Haplotype	Clone	Results from Protein and cDNA Sequence Comparisons
C14	d	pH-2II	Not K^d or D^d
		pH-2I	Different from clone pH-2II
RDM-4	k	pH-2III	Not K^k

The predicted amino acid sequence of clone pH-2II corresponds in 19 out of 19 positions that can be compared (Nairn et al., 1980) to the L^d amino acid sequence. It is therefore possible that pH-2II encodes the L^d molecule.

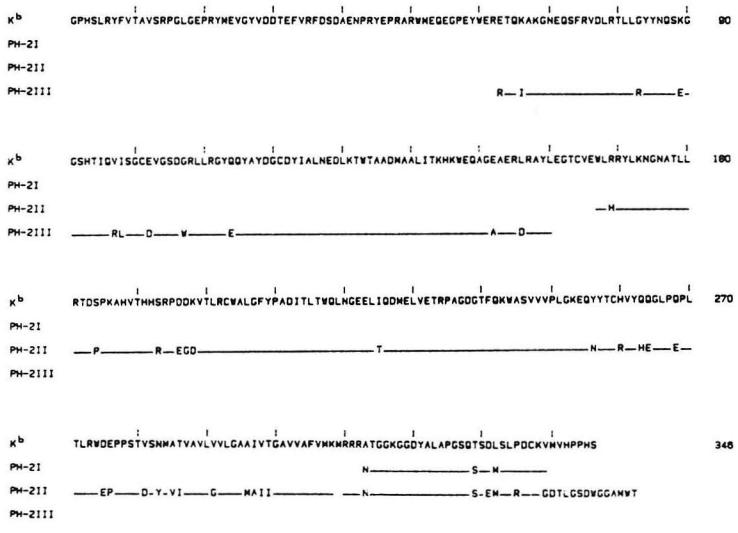

Figure 3. A Comparison of the Amino Acid Sequence of the Mouse K^b Molecule and the Translated Protein Sequences of the Three H-2 cDNA Clones

The mouse H-2K^b sequence has been published by Martinko et al. (1980) and Uehara et al. (1981). A gap of one amino acid at position 309 has been inserted into the predicted amino acid sequence of clone pH-2II to achieve maximum homology to the K^b molecule. The single-letter code has been used for amino acids: A, Ala; B, Asp or Asn; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; J, Glu or Gln; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln, R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

Solid line: identity to the K^b sequence.

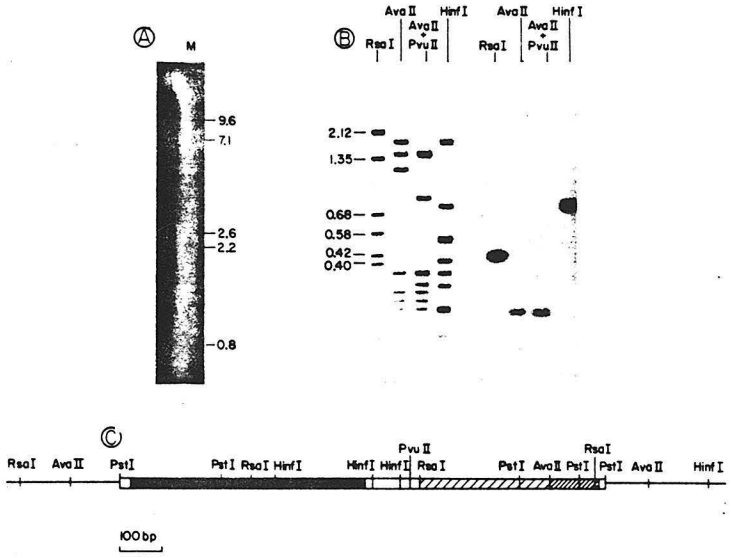

Figure 4. Localization of Repetitive DNA Sequences on the 3′ Untranslated Region of Clone pH-2II

(A) The fragments from 10 μg of Eco RI–cleaved BALB/c mouse liver DNA (5×10^{-18} mole) were separated on a 0.6% agarose gel, transferred to a nitrocellulose filter and hybridized to clone pH-2II. The stringent wash was in 0.1 × SSC, 0.1% SDS at 68°C. Restriction fragments (7.5×10^{-18} mole per discrete fragment) of a pBR322 derivative were run in parallel (lane M) and served as molecular weight markers. Sizes are given in kb.

(B) The ethidium bromide–stained agarose gel (2%) on the left shows the restriction fragments obtained from the plasmid clone pH-2II upon digestion with the indicated nucleases. Fragment sizes are in kb. After blotting and hybridization with total mouse liver DNA, the autoradiogram on the right was obtained. Only sequences that are repetitive in mouse DNA are expected to hybridize under the conditions employed (Steinmetz et al., 1980).

(C) The map of the insert in clone pH-2II shows the localization of a moderately repetitive sequence between restriction sites for Rsa I and Ava II and a highly repetitive sequence between Ava II and the 3′ end on the 3′ untranslated region. Hatching: regions containing repetitive sequences; open boxes: G–C tails; filled boxes: coding region; thin line: pBR322 flanking sequences.

indicating the presence of a moderately repetitive sequence in the 3′ untranslated region between the Rsa I and the Ava II site and a highly repetitive sequence between the Ava II site and the 3′ end (Figure 4C). Sequence comparisons (not shown) indicate that these repetitive sequences do not fall into the Alu family described by Jelinek et al. (1980). To our knowledge this is the first example of repetitive sequence elements on the 3′ untranslated region of a messenger RNA encoding a characterized protein.

The Genes Encoding Mouse Transplantation Antigens Do Show Convincing Homologies with Immunoglobulin Genes

Several features of transplantation antigens suggest that they may be evolutionarily related to immunoglobulins. First, the β_2–microglobulin polypeptide associated with transplantation antigens is ~30% homologous to immunoglobulin constant region domains, and accordingly has been denoted a "free immunoglobulin domain" (Peterson et al., 1972). Since immunoglobulin domains of light and heavy chains interact, perhaps β_2–microglobulin and the transplantation antigens interact through similar immunoglobulin-like domains. Second, immunoglobulins are comprised of multiple homology units or domains. These homology units, about 110 residues in length, are characterized by a centrally placed disulfide bridge spanning about 60 residues. The transplantation antigen also has two disulfide bridges, each spanning about 60 residues—that is, residues 101 to 164 and 203 to 259. Moreover, in human transplantation antigens, an 89 residue stretch of amino acid sequence including the second disulfide bridge (~181–270) shows a statistically significant homology (35%) to portions of immunoglobulin constant region domains at the protein level (Strominger et al., 1980). However, six sequence gaps must be inserted into the two sequences to obtain a 35% homology. One of the interesting questions that arises from these data is whether the protein homology arose by divergent (genes diverging from a common ancestral gene) or convergent (two independent genes converging toward a common protein sequence) evolution. This question cannot be answered at the protein level.

We used two computer programs to determine whether the genes for transplantation antigens were homologous to immunoglobulin genes. Initial analyses of the homology relationships of the H-2 cDNA clones to one another and to various immunoglobulin genes were done with the dot matrix computer program. Because this program cannot detect more distant and dispersed evolutionary relationships, we developed a second computer program, the best-fit matrix program (see Experimental Procedures for a description of these programs).

A best-fit matrix analysis of the pH-2II clone against the mouse immunoglobulin $C_\mu 4$ domain is given in Figure 5. The extended diagonal lines ("a" or "b") denote two homologous sequences located between nucleotides 119–364 of clone pH-2II and 1549–1809 in the $C_\mu 4$ domain of the μ gene. The homology is demonstrated at the DNA and protein levels in Figure 6. The two sequences are 51% (122 out of 237) homologous at the nucleotide level after the placement of two sequence gaps into each sequence. When

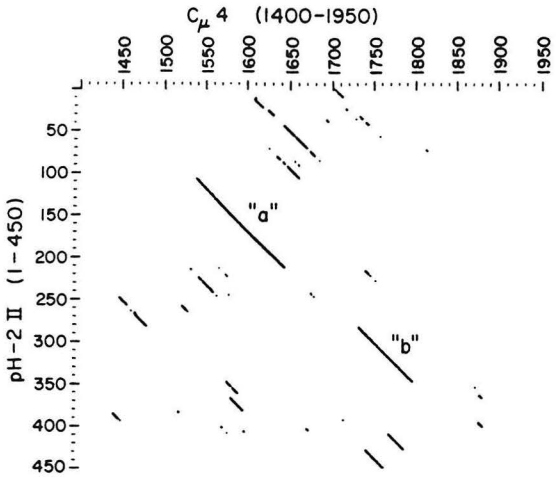

Figure 5. Homology between the pH-2II cDNA Sequence and the $C_\mu 4$ Domain of the Mouse Immunoglobulin μ Chain as Shown by the Best-fit Matrix Routine

We compared the pH-2II cDNA sequence from position 1 to 450 (vertical axis) with the DNA sequence of the $C_\mu 4$ domain from position 1400 to 1950 (horizontal axis) using the best-fit matrix program as described in Experimental Procedures. The numbering for the pH-2II sequence is the same as in Figure 2. The mouse μ chain sequence is from Kawakami et al. (1980) and nucleotide numbering for the μ gene is as published. Homologous regions in the analyzed sequences appear as a $-45°$ line in the matrix. The straight lines denoted by "a" and "b" indicate regions of significant homology between H-2 and the $C_\mu 4$ domain (see Figure 6). The significance of the homology is reinforced by the fact that the same regions of the pH-2II cDNA sequence are homologous with the $C_\mu 1$, $C_\mu 2$ and $C_\mu 3$ exons (results not shown).

the best-fit matrix is used to compare the same stretch of pH-2II sequence against the $C_\mu 1$, $C_\mu 2$ and $C_\mu 3$ domains, similar, although somewhat less extensive, homologies are noted (data not shown). Indeed, the same is true for the $C_{\gamma 1}$ and $C_{\gamma 2b}$ domains (data not shown). Moreover, this sequence in the pH-2II clone appears, if anything, to be more closely related to the C_μ domains than the C_μ domains are to one another. The extended diagonal line "a" in Figure 5 represents a 72 base sequence (119–190) from clone pH–2II that is strikingly homologous to corresponding regions in the immunoglobulin domains $C_\mu 1$, $C_\mu 2$ and $C_\mu 4$ and is somewhat less homologous to that of the $C_\mu 3$ domain—53%, 52%, 60% and 36%, respectively (Figure 7). Within this 72 base stretch, there are 21 base positions conserved between the $C_\mu 1$, $C_\mu 2$ and $C_\mu 4$ sequences. Twenty of these 21 are conserved in clone pH-2II as well. This homology has been determined without placing any sequence gaps—thus permitting us to ask how frequently homologies at these levels would be seen if every possible stretch of 72 nucleotides in the pH-2II clone were compared against all possible blocks of 72 nucleotides in the mouse C_μ gene. About 1.3×10^6 such comparisons were made; the mean homology for these comparisons is 25% with a standard deviation of 5.4%. Thus the homologies exhibited by the comparisons of the pH-2II clone against the $C_\mu 1$, $C_\mu 2$, $C_\mu 3$ and $C_\mu 4$ domains in Figure 7 fell 5.2, 5.0, 2.0 and 6.6 standard deviations from the mean and, accordingly, are all highly statistically significant.

It is important to point out that the third base positions in codons are highly conserved in the comparisons shown in Figures 6 and 7. Indeed, 45% of the third base positions in the comparison between clone pH-2II and the $C_\mu 4$ domain in Figure 6 are conserved. Moreover, six third base positions are absolutely conserved in the 72 nucleotide sequence for pH-2II and the $C_\mu 1$, $C_\mu 2$ and $C_\mu 4$ domains (Figure 7), even though in three of these cases the amino acid is not conserved. These observations strongly suggest that the homologies between transplantation antigens and immunoglobulin arose by divergent rather than convergent evolution, because convergent evolution drives different genes to produce similar protein sequences without any selective pressures for the conservation of the third base positions in codons.

The homologous stretch of sequence in the transplantation antigens is located on the extracellular part of the molecule proximal to the cell membrane. Perhaps the reason this sequence is conserved is that this portion of the transplantation antigen must fold into the classic "antibody fold" (Poljak et al., 1973) in order to interact effectively in a noncovalent manner with β_2–microglobulin, which is folded into a similar configuration—much as the light and heavy chain domains of immunoglobulins interact with one another through a molecular complementarity in their similar antibody folds.

DNA Blots on Mouse Liver DNA Suggest That Transplantation Genes Are a Multigene Family

The homology between immunoglobulins, which are encoded by several large multigene families, and transplantation antigens raises questions about the multiplicity of genes encoding transplantation antigens. Indeed, several investigators have considered the possibility that the mouse genome contains many more transplantation antigen genes than required for the three or four polypeptides that are expressed in individual mice (Bodmer, 1973; Silver and Hood, 1976). In addition, there appear to be several other loci on chromosome 17 that encode gene products possibly related to transplantation antigens because of their similar size and association with β_2–microglobulin (for examples, the T cell differentiation antigens, such as TL, Qa1 and Qa2). Protein sequence data, however, are lacking for the T cell differentiation antigens, so that the degree of their sequence relatedness to transplantation antigens is unknown.

To obtain a rough estimate of the number of DNA sequences in the mouse genome that crosshybridize

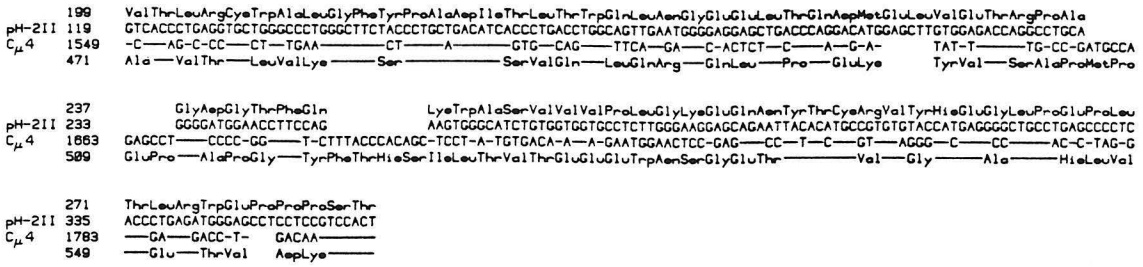

Figure 6. A Comparison between the DNA Sequences of Clone pH-2II and the Fourth Constant Region Domain of the Mouse Immunoglobulin μ Heavy Chain

A comparison is made between the pH-2II cDNA sequence from nucleotide positions 119 to 364 (Figure 2) and the fourth constant region exon ($C_\mu 4$) of the μ gene from positions 1549 to 1809. This comparison includes the two homology regions identified in Figure 5 and shows that an overall homology of 51% can be achieved for the whole region with only two gaps in each sequence. The amino acid sequences encoded by the pH-2II and the $C_\mu 4$ DNA sequences are given above and below the nucleotide sequences, respectively.
Solid line: $C_\mu 4$ nucleotides and amino acids that are identical to the pH-2II sequence.

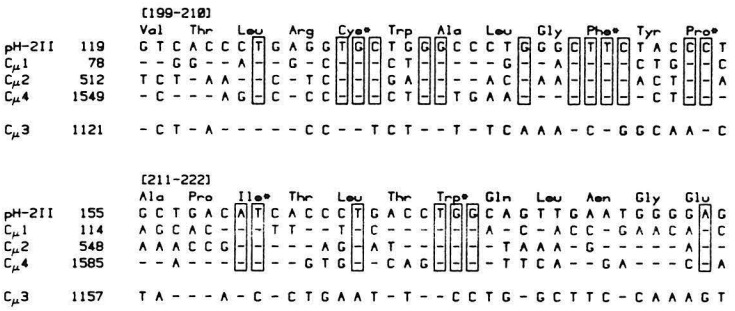

Figure 7. A Comparison between the DNA Sequences of the "a" Region of the pH-2II cDNA Clone and the Homologous Portions for the Immunoglobulin $C_\mu 1$, $C_\mu 2$, $C_\mu 3$ and $C_\mu 4$ Domains

The DNA sequence of clone pH-2II from nucleotide positions 119 to 190 (Figure 2) is compared with the DNA sequences of the four μ chain domains $C_\mu 1$ to $C_\mu 4$ (Kawakami et al., 1980) found to be homologous by the best-fit matrix comparison (see Figure 5). Dashes: identical nucleotides to the pH-2II sequence. Boxes: sequences that are conserved between pH-2II and the $C_\mu 1$, $C_\mu 2$ and $C_\mu 4$ domains. The amino acid sequence encoded by pH-2II (positions 199 to 222) is given above the DNA sequence. Conserved amino acids are marked by an asterisk.

with our cDNA probes for transplantation antigens, we carried out Southern blot analyses of various mouse DNAs. Figure 8A shows a DNA blot analysis of a Bam HI digestion of BALB/c liver DNA with the pH-2II subclone. Figure 8B shows similar Bam HI analyses of sperm, embryo and liver DNA from BALB/c mice (d haplotype) and of liver DNAs from mice of the b, k and the recombinant a (K^kD^d) haplotypes, with the clone pH-2III as a probe. Several observations can be made. First, with both probes multiple bands of different intensities are detected. About 15 bands are identified with the pH-2II probe and about 12 with pH-2III probe. Second, the pH-2III and pH-2II probes, which represent the 5′ and 3′ areas of the coding sequence for transplantation antigens, hybridize to different, although probably overlapping, sets of restriction fragments. An analysis of five genomic clones crossreacting with these cDNA probes suggests that each putative transplantation antigen gene is cleaved a single time with the Bam HI enzyme and that in each case the pH-2II or pH-2III probes hybridize to just a single fragment (K. Moore, unpublished data).

The finding of multiple bands can be explained in several ways. First, the genes encoding transplantation antigens may be divided by many introns and thus many of the bands may represent a single gene. However, this explanation appears unlikely for many of the bands because of the detailed restriction analysis of the five distinct genomic clones mentioned above. Second, many of the bands may represent different genes, encoding transplantation antigens, as well as nonfunctional pseudogenes. Indeed, two factors may lead us to underestimate the number of distinct genes. Since the bands differ markedly in intensity, perhaps several of the heavy bands represent multiple distinct genes. In addition, our cDNA probes may not crossreact with all of the genes encoding transplantation antigens. It is unlikely, however, that there will be many more genes (bands), because variation of the stringency of hybridization does not alter the multiplicity of the bands. Accordingly, if all of these bands represent transplantation antigen genes, there could be 15 or more distinct genes. Third, perhaps the T cell differentiation antigens mentioned earlier are sufficiently homologous with transplantation antigens to crossreact with the cDNA probes employed. A detailed analysis of corresponding genomic clones and expressed proteins should allow us to differentiate between the latter two possibilities. We conclude that the genes encoding the transplantation antigens constitute a multigene family.

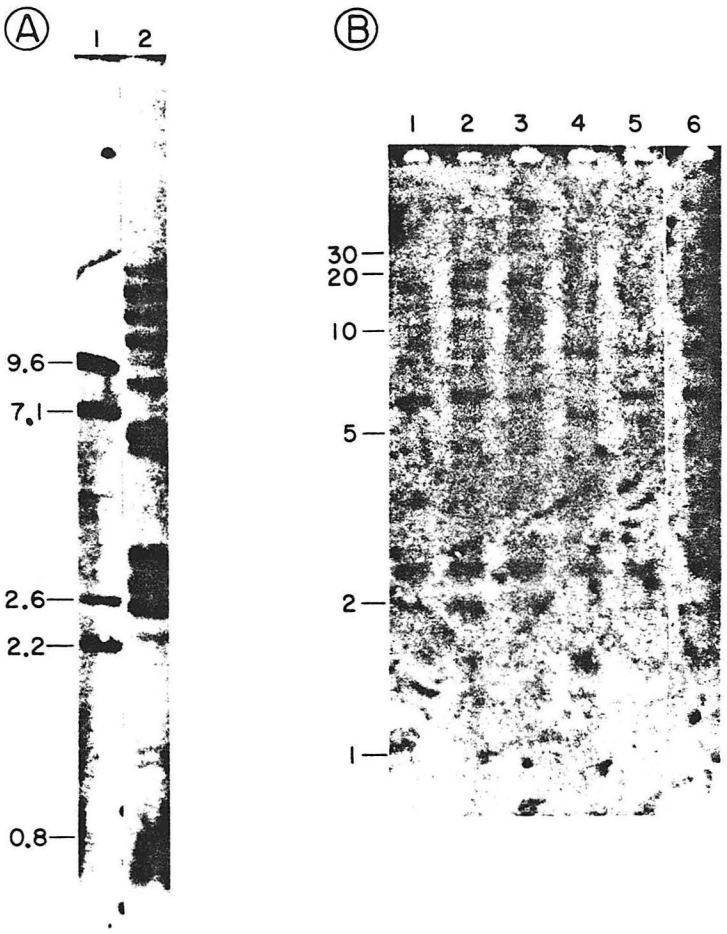

Figure 8. Southern Blot Hybridization of Mouse DNA with the pH-2II Subclone and the pH-2III cDNA Clone as Probes

(A) (Lane 2) 10 μg (5.6×10^{-18} mole) of BALB/c mouse liver DNA were cleaved with Bam HI to completion, separated on a 0.6% agarose gel and transferred to a nitrocellulose filter. Hybridization was with 50 ng/ml of a subclone containing the 688 bp Pst I–Pvn II fragment of pH-2II in pBR322 (see Figure 4). The final wash was in 0.1 × SSC, 0.1% SDS at 65°C. The hybridization markers in lane 1 are the same as in Figure 4, except that 5×10^{-18} mole per discrete fragment were used. (B) Each lane contained 3 μg of Bam HI–digested mouse DNA. Hybridization was carried out with 20 ng of the 287 bp Pst I fragment of clone pH-2III, which had been ligated to itself and labeled by nick translation. The final wash was in 1 × SSC, 0.1% SDS at 68°C. (Lanes 1, 2 and 3) sperm, embryo and liver DNA from BALB/c mouse (d haplotype); (lanes 4, 5 and 6) liver DNA from CBA (k haplotype), A/J (a haplotype) and B10 mice (b haplotype). λ DNA fragments were run in parallel and served as molecular weight markers. Sizes are in kb.

One final observation is that a single restriction enzyme polymorphism is seen between the haplotypes tested. This polymorphism occurs in the k haplotype, where a strongly hybridizing 5.4 kb band is seen instead of the 6.2 kb band found in the d haplotype (Figure 8B). Moreover, in the DNA from the recombinant a haplotype (K^kD^d) one finds the 6.2 kb band. This indicates that the polymorphism is encoded by the D region of the H-2 complex. In view of the extensive amino acid differences noted among the homologous transplantation antigens from different haplotypes (~5–10%), it is surprising to find only one restriction enzyme polymorphism. The simple interpretation of these observations is that the noncoding sequences of transplantation antigens are far more highly conserved than their coding regions.

DNA Blots on Sperm DNA Suggest That There Is No Rearrangement of the Genes Encoding Transplantation Antigens during the Differentiation of Murine Somatic Cells

The homology between the genes encoding immunoglobulins and transplantation antigens raises the interesting possibility that DNA rearrangements may be correlated with the expression of transplantation antigens, as they are associated with the expression of immunoglobulins (Brack et al., 1978; Seidman et al., 1979; Early et al., 1980). Transplantation antigens are not endogenously expressed by sperm cells (undifferentiated DNA), whereas they are expressed by virtually all somatic cells (differentiated DNA) (Klein, 1975). It is important to remember that at least four transplantation antigens are expressed in the BALB/c mouse, and if rearrangements were required for expression, all four genes should rearrange. By Southern blot hybridization, we analyzed Eco RI–, Bam HI– and Hind III–digested DNA from BALB/c sperm, embryo and liver DNA with the pH-2III clone and the single-copy subclone of pH-2II. Figure 8B shows the results obtained for the Bam digestion of these DNAs probed with the clone pH-2III. In all cases, no difference between differentiated and undifferentiated DNA was seen (Figure 8B lanes 1–3). Our failure to find differences between sperm and differentiated DNA by this method suggests that DNA rearrangements are unlikely within the gene segments of

transplantation antigens during differentiation. However, this is not unequivocal proof against DNA rearrangements, since all three restriction enzymes may cut between the putative recombination site and the portions of the gene for which we have a probe—a possibility rendered less likely by the fact there are at least four genes for transplantation antigens expressed in BALB/c mice.

The availability of cDNA clones for transplantation antigens will now allow the detailed study of the H-2 complex of chromosome 17 in the mouse, and will help to answer questions concerning the organization, expression and evolution of these H-2 genes.

Experimental Procedures

Materials

Restriction nucleases, T4 DNA kinase, exonuclease III and E. coli DNA polymerase I large fragment were purchased from New England Biolabs. Calf liver tRNA was obtained from Boehringer. Reverse transcriptase and E. coli DNA polymerase I, used for double-stranded cDNA synthesis, were gifts from J. Beard and M. Bond, respectively. Mouse sperm DNA was obtained from I. Weissman. The E. coli strain MC1061, originally provided by M. Casadaban, was obtained from T. Sargent. The mouse lymphoma cell line RDM-4 was provided by M. F. Mescher.

Preparation of Poly(A)$^+$ RNA

RDM-4 (H-2^k) and C14 (H-2^d) tumor cells were grown in Dulbecco's Modified Eagle's medium supplemented with 10% heat-inactivated fetal calf serum, 10 mM nonessential amino acids, 0.1 mg/ml glutamine, 50 μg/ml penicillin and 50 μg/ml streptomycin to a density of 4×10^6 and 2×10^6 cells/ml, respectively. Usually 10^9 cells were harvested, pelleted for 10 min at 1000 × g at 0°C and washed once in 50 ml 150 mM NaCl, 40 mM phosphate, pH 7.2. RNA was then isolated by precipitation of polysomes (Efstratiadis and Kafatos, 1976), or by the guanidinium thiocyanate procedure (Chirgwin et al., 1979) as described by Fyrberg et al. (1980). Poly(A)$^+$ RNA was selected by a single passage over oligo(dT)–cellulose (Efstratiadis and Kafatos, 1976).

Cloning of Double-stranded cDNA

Double-stranded cDNA was synthesized following published procedures (Friedman and Rosbash, 1977; Buell et al., 1978; Chang et al., 1978; Wickens et al., 1978; Efstratiadis and Villa-Komaroff, 1979) with minor modifications. About 2 μg of double-stranded cDNA was obtained from 10 μg of poly(A)$^+$-selected mRNA (estimated to be contaminated with poly(A)$^-$ RNA to about 50%). Its single-stranded length, as determined by alkaline agarose gel electrophoresis, was on the average 600 nucleotides. No size selection was employed. pBR322 DNA was cleaved with Pst I and the linearized molecule was purified by electrophoresis on a 1% agarose gel. It was recovered by electroelution in a dialysis bag (Smith, 1980) and was purified by BD-cellulose chromatography. About eight dGMP residues were added per 3′ end (Roychoudhury and Wu, 1980). To the double-stranded cDNA about 10–20 dCMP residues were added per 3′ end.

Equimolar amounts of tailed cDNA and tailed vector DNA were combined in 0.1 M NaCl, 10 mM Tris, 0.25 mM EDTA, pH 7.5 at a final DNA concentration of 1 μg/ml, heated to 64°C for 8 min and annealed by incubation in a water bath at 42–43°C. After 2 hr the water bath was switched off and cooled to room temperature overnight. Five micrograms of tRNA were then added as a carrier and the DNA was precipitated with ethanol. The pellet was washed with 80% ethanol, dried, dissolved in 0.1 M NaCl, 10 mM morpholinopropane sulfonic acid, pH 7.0 (10 μl per 300 ng of DNA) and stored on ice.

MC1061 E. coli cells were transformed following the procedure described by Kushner (1978), except that 150 ng of DNA were used per discrete transformation and L broth was used instead of Z broth. After transformation the cells were pelleted, resuspended in 0.2 ml of L broth and spread on nitrocellulose filters on plates containing 15 μg/ml tetracycline at a density of about 1000 colonies per filter. For both the C14 and RDM-4 libraries, 6000–8000 tetracycline-resistant colonies were obtained from 50 ng of double-stranded cDNA. Replica plating, amplification, storage and lysis of colonies on nitrocellulose filters were essentially as described by Hanahan and Meselson (1980).

Colony hybridization with a purified Pst I fragment containing a cloned HLA cDNA (Sood et al., 1981) was carried out in 3 × SSC, 1× Denhardt's solution (Denhardt, 1966), 10% dextran sulfate, 0.1% SDS, poly(rC), poly(rA), poly(rG), for 1 hr at 65°C and then at 55°C overnight. Filters were washed at 55°C in 3 × SSC, 1× Denhardt's solution, and then in 1 × SSC.

DNA Sequence Analysis

Restriction fragments were labeled at their 5′ ends with polynucleotide kinase and γ-^{32}P-ATP or at their 3′ ends with E. coli DNA polymerase I large fragment and α-^{32}P-dNTPs. The 3′ ends of Rsa I fragments were labeled by successive treatment with exonuclease III and E. coli DNA polymerase I large fragment, similar to the procedure described by Smith and Calvo (1980). The DNA was strand-separated and single strands were sequenced by the chemical degradation procedure (Maxam and Gilbert, 1980). For the G+A reaction the procedure described by Gray et al. (1978) was used. Reaction products were electrophoresed on 40 cm 20% acrylamide sequencing gels and 80 cm 5% and 8% acrylamide sequencing gels (Smith and Calvo, 1980).

DNA Blot Hybridization

DNA blots were prepared and hybridized with nick-translated probes as described (Schnell et al., 1980), except that the hybridization solution was supplemented with poly(rA), poly(rG) and poly(rC) (10 μg/ml each).

Computer Homology Programs

Initial sequence analysis of the H-2 cDNA clones was done with the aid of a dot matrix computer program, which belongs to DNAMST, a database and analysis system under development in this laboratory. This routine, presumably similar to those described by other laboratories (Efstratiadis et al., 1980; Hieter et al., 1980), looks at all possible alignments between all possible DNA sequence fragments of a desired length that can be generated from two sequences. A positive homology is scored if the percent homology of a comparison equals or exceeds a preset minimum. Homology within these parameters is displayed graphically as a dot. The coordinates of the dots are equivalent to the base positions of the lead bases of the two compared DNA segments on an X–Y field with axes represented by the two whole sequences. Regions of homology appear as various discrete patterns, but primarily as lines parallel to the diagonal of the matrix.

The H-2 clones were compared in this manner (data not shown) both with themselves and with each other, as well as against the mouse genomic C_μ, $C_{\gamma 1}$ and $C_{\gamma 2b}$ genes. The 3′ flanking sequence also was compared to representative Alu family sequences.

Dot matrix–like routines have a serious limitation when used to compare sequences that are distantly related, and that may have significant homology that is distributed in a diffuse manner. Often the dot matrix program will increase the background to the point that real homology is obscured, if the length of the unit sequence compared or the percentage of homology required for a positive score is lowered. To get around this limitation, we used another routine that is graphically similar to the dot matrix, but differs in how it scores and evaluates homology between DNA fragments. The best-fit matrix routine establishes a homology score for each possible single base alignment between any two sequences. This score depends on the positive or negative score of the two base positions in question, as well as on that between the corresponding bases of the 5′ and 3′ neighboring sequences. The contribution to a given score by any flanking se-

quence homology is inversely proportional to its distance in sequence length from the aligned base pair. Each base of one DNA sequence is compared in this manner with all the base positions of the other DNA sequence and a score for each comparison is generated. A dot is then plotted, as it is with the dot matrix routine, to represent homology, but only at those matrix coordinates that equal the minimum score generated for a given base.

The results of this type of plot resemble a dot matrix with diagonally parallel lines representing regions of relative homology. When the pH-2II clone was compared to the C_μ gene sequence in this manner, two lines had high homology scores ("a" and "b" in Figure 5) and the remainder of the shorter lines had lower scores. A detailed discussion of these programs is now in preparation (T. Hunkapiller et al., unpublished data).

Acknowledgments

We thank Michael Douglas for computer graphics. M. S. is the recipient of a fellowship from the Deutsche Forschungsgemeinschaft and D. F. and T. H. are NIH trainees. J. G. F. is supported by an NIH postdoctoral fellowship. This work was supported by grants from the National Institutes of Health.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received December 24, 1980; revised January 23, 1981

References

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

Brack, C., Hirama, M., Lenhard-Schuller, R. and Tonegawa, S. (1978). A complete immunoglobulin gene is created by somatic recombination. Cell *15*, 1–14.

Buell, G. N., Wickens, M. P., Payvar, F. and Schimke, R. T. (1978). Synthesis of full length cDNAs from four partially purified oviduct mRNAs. J. Biol. Chem. *253*, 2471–2482.

Chang, A. C. Y., Nunberg, J. H., Kaufman, R. J., Erlich, H. A., Schimke, R. T. and Cohen, S. N. (1978). Phenotypic expression in E. coli of a DNA sequence coding for mouse dihydrofolate reductase. Nature *275*, 617–624.

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. and Rutter, W. J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry *18*, 5294–5299.

Coligan, J. E., Kindt, T. J., Nairn, R., Nathenson, S. G., Sachs, D. H. and Hansen, T. H. (1980). Primary structural studies of an H-2L molecule confirm that it is a unique gene product with homology to H-2K and H-2D antigens. Proc. Nat. Acad. Sci. USA *77*, 1134–1138.

Denhardt, D. T. (1966). A membrane-filter technique for the detection of complementary DNA. Biochem. Biophys. Res. Commun. *23*, 641–646.

Early, P., Huang, H., Davis, M., Calame, K. and Hood, L. (1980). An immunoglobulin heavy chain variable region gene is generated from three segments of DNA, V_H, D and J_H. Cell *19*, 981–992.

Efstratiadis, A. and Kafatos, F. C. (1976). The chorion of insects: techniques and perspectives. In Methods in Molecular Biology, *8*, J. A. Last, ed. (New York: Dekker), pp. 1–124.

Efstratiadis, A. and Villa-Komaroff, L. (1979). Cloning of double-stranded cDNA. In Genetic Engineering: Principles and Methods, *1*, J. K. Setlow and A. Hollaender, eds. (New York: Plenum), pp. 15–36.

Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C. and Proudfoot, N. J. (1980). The structure and evolution of the human β-globin gene family. Cell *21*, 653–668.

Friedman, E. Y. and Rosbash, M. (1977). The synthesis of high yields of full-length reverse transcripts of globin mRNA. Nucl. Acids Res. *4*, 3455–3471.

Fyrberg, E. A., Kindle, K. L., Davidson, N. and Sodja, A. (1980). The actin genes of Drosophila: a dispersed multigene family. Cell *19*, 365–378.

Gorer, P. A. (1938). The antigenic basis of tumour transplantation. J. Pathol. Bacteriol. *47*, 231–252.

Gorer, P. A., Lyman, S. and Snell, G. D. (1948). Studies on the genetic and antigenic basis of tumour transplantation: linkage between a histocompatibility gene and "fused" in mice. Proc. R. Soc. (Lond.) B *135*, 499–505.

Gray, C. P., Sommer, R., Polke, L., Beck, E. and Schaller, H. (1978). Structure of the origin of DNA replication of bacteriophage fd. Proc. Nat. Acad. Sci. USA *75*, 50–53.

Hanahan, D. and Meselson, M. (1980). Plasmid screening at high colony density. Gene *10*, 63–67.

Hansen, T. H., Ozato, K., Melino, M. R., Coligan, J. E., Kindt, T. J., Jandinski, J. J. and Sachs, D. H. (1981). A gene cluster in the *H-2D* region: evidence in two haplotypes for at least three *D* region-encoded molecules, D, L, and R. J. Immunol., in press.

Herrmann, S. H. and Mescher, M. F. (1979). Purification of the H-$2K^k$ molecule of the murine major histocompatibility complex. J. Biol. Chem. *254*, 8713–8716.

Hieter, P. A., Max, E. E., Seidman, J. G., Maizel, J. V., Jr. and Leder, P. (1980). Cloned human and mouse kappa immunoglobulin constant and J region genes conserve homology in functional segments. Cell *22*, 197–107.

Jelinek, W. R., Toomey, T. P., Leinwand, L., Duncan, C. H., Biron, P. A., Choudary, P. V., Weissman, S. M., Rubin, C. M., Houck, C. M., Deininger, P. L. and Schmid, C. W. (1980). Ubiquitous, interspersed repeated sequences in mammalian genomes. Proc. Nat. Acad. Sci. USA *77*, 1398–1402.

Kawakami, T., Takahashi, N. and Honjo, T. (1980). Complete nucleotide sequence of mouse immunoglobulin μ gene and comparison with other immunoglobulin heavy chain genes. Nucl. Acids Res. *8*, 3933–3945.

Klein, J. (1975). Biology of the Mouse Histocompatibility-2 Complex. (Berlin: Springer-Verlag).

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

Kushner, S. (1978). An improved method for transformation of Escherichia coli with Col E1 derived plasmids. In Genetic Engineering, H. W. Boyer and S. Nicosia, eds. (Amsterdam: Elsevier/North Holland Biomedical Press), pp. 17–23.

Maizels, R. M., Frelinger, J. A. and Hood, L. (1978). Partial amino acid sequences of mouse transplantation antigens. Immunogenetics *7*, 425–444.

Martinko, J. M., Uehara, H., Ewenstein, B. M., Kindt, T. J., Coligan, J. E. and Nathenson, S. G. (1980). Primary structure of murine major histocompatibility complex alloantigens: completion of the sequence of the amino-terminal 284 residues of H-$2K^b$. Biochemistry *19*, 6188–6193.

Maxam, A. M. and Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. Meth. Enzymol. *65*, 499–560.

Nairn, R., Nathenson, S. G. and Coligan, J. E. (1980). Isolation, characterization and amino acid sequence studies of the cyanogen bromide fragments of the H-$2D^d$ glycoprotein. Eur. J. Immunol. *10*, 495–503.

Nathenson, S. G., Uehara, H., Ewenstein, B. M., Kindt, T. J. and Coligan, J. E. (1981). Primary structural analysis of transplantation antigens of the murine H-2 major histocompatibility complex. Ann. Rev. Biochem., in press.

Orr, H. T., Lancet, D., Robb, R. J., Lopez de Castro, J. A. and Strominger, J. L. (1979). The heavy chain of human histocompatibility antigen HLA-B7 contains an immunoglobulin-like region. Nature *232*, 266–270.

Peterson, P. A., Cunningham, B. A., Berggard, I. and Edelman, G. M. (1972). β_2-Microglobulin—a free immunoglobulin domain. Proc. Nat. Acad. Sci. USA *69*, 1697–1701.

Ploegh, H. L., Orr, H. T. and Strominger, J. L. (1980). Molecular cloning of a human histocompatibility antigen cDNA fragment. Proc. Nat. Acad. Sci. USA *77*, 6081–6085.

Poljak, R. J., Amzel, L. M., Avey, H. P., Chen, B. L., Phizackerley, R. P. and Saul, F. (1973). Three-dimensional structure of the Fab' fragment of a human immunoglobulin at 2.8 Å resolution. Proc. Nat. Acad. Sci. USA *70*, 3305–3310.

Rothbard, J. B., Hopp, T. P., Edelman, G. M. and Cunningham, B. A. (1980). Structure of the heavy chain of the H-2K^k histocompatibility antigen. Proc. Nat. Acad. Sci. USA *77*, 4239–4243.

Roychoudhury, R. and Wu, R. (1980). Terminal transferase-catalyzed addition of nucleotides to the 3′ termini of DNA. Meth. Enzymol. *65*, 43–62.

Schnell, H., Steinmetz, M., Zachau, H. G. and Schechter, I. (1980). An unusual translocation of immunoglobulin gene segments in variants of the mouse myeloma MPC11. Nature *286*, 170–173.

Seidman, J. G., Max, E. E. and Leder, P. (1979). A K-immunoglobulin gene is formed by site specific recombination without further somatic mutation. Nature *280*, 370–375.

Shearer, G. M. and Schmitt-Verhulst, A. M. (1977). Major histocompatibility complex restricted cell-mediated immunity. Adv. Immunol. *25*, 55–91.

Silver, J. and Hood, L. (1976). Preliminary amino acid sequences of transplantation antigens: genetic and evolutionary implications. In Contemporary Topics in Molecular Immunology, *5*, H. N. Eisen and R. A. Reisfeld, eds. (New York: Plenum Publishing Corporation), pp. 35–68.

Smith, D. R. and Calvo, J. M. (1980). Nucleotide sequence of the E. coli gene coding for dihydrofolate reductase. Nucl. Acids Res. *8*, 2255–2274.

Smith, H. O. (1980). Recovery of DNA from gels. Meth. Enzymol. *65*, 371–380.

Snell, G. D., Dausset, J. and Nathenson, S. (1976). Histocompatibility. (New York: Academic Press).

Sood, A. K., Pereira, D. and Weissman, S. M. (1981). Isolation and partial nucleotide sequence of a cDNA clone for human histocompatibility antigen HLA-B by use of an oligodeoxynucleotide primer. Proc. Nat. Acad. Sci. USA, in press.

Steinmetz, M., Höchtl, J., Schnell, H., Gebhard, W. and Zachau, H. G. (1980). Cloning of V region fragments from mouse liver DNA and localization of repetitive DNA sequences in the vicinity of immunoglobulin gene segments. Nucl. Acids Res. *8*, 1721–1729.

Strominger, J. L., Orr, H. T., Parham, P., Ploegh, H. L., Mann, D. L., Bilofsky, H., Saroff, H. A., Wu, T. T. and Kabat, E. A. (1980). An evaluation of the significance of amino acid sequence homologies in human histocompatibility antigens (HLA-A and HLA-B) with immunoglobulins and other proteins, using relatively short sequences. Scand. J. Immunol. *11*, 573–592.

Uehara, H., Coligan, J. E. and Nathenson, S. G. (1981). Isolation and sequence analysis of membrane and hydrophilic segments of the H-2K^b alloantigen. Biochemistry, in press.

Vitetta, E. S. and Capra, J. O. (1978). The protein products of the murine 17th chromosome: genetics and structure. Adv. Immunol. *26*, 147–193.

Wickens, M. P., Buell, G. N. and Schimke, R. T. (1978). Synthesis of double-stranded DNA complementary to lysozyme, ovomucoid and ovalbumin mRNAs. J. Biol. Chem. *253*, 2483–2495.

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

Chapter 2

METHODS FOR CLONING AND SEQUENCING IN BACTERIOPHAGE M13

Introduction

The following are the methods that I have used to clone and sequence in the filamentous bacteriophage M13. For the sake of completeness, most of the useful protocols for handling M13 phage are included, although many of these do not differ significantly from versions to be found elsewhere (1-12). The sections on low melting point agarose and deletion subcloning are new. The latter replaces the random shotgun method as the way of generating overlapping sequencing subclones.

Vectors: Bacteriophage M13 exists in two forms during its life cycle in its host, *E. coli*. The virion contains single-stranded DNA while the intracellular replicative form (RF) is a double-stranded, closed circle. The replicative form is useful for carrying out cloning manipulations (endonuclease digestions, ligations, etc.) while the single-stranded form is a convenient template for Sanger/dideoxy DNA sequencing (1). The M13mp-series vectors have a universal oligonucleotide priming site for sequencing located immediately adjacent to a polylinker (Figure 1). The polylinker contains numerous useful restriction endonuclease cleavage sites unique to the phage genome.

The genome of the M13mp phage contains part of the *E. coli* lac operon, as well as the wild type M13 genome. Both the primer binding site and the polylinker are located in the β-galactosidase gene of the lac operon, and growth of mp-series phage on an IPTG- and Xgal-containing plate result in blue plaques. Inserts cloned at the polylinker inactivate the β-galactosidase gene and result in phage that form colorless plaques on indicator plates.

Hosts: *E. coli* hosts JM101 and JM103 can be used with equal results. Only male bacteria (those containing the F episome) can be infected by M13 phage. JM101 and JM103 are pro^- bacteria containing a complementing pro^+ gene on the F episome. Growth in the absence of proline, on a minimal agar plate, selects for

male bacteria. For this reason, JM101/103 are always streaked out on minimal plates, and the resulting colonies used to start cultures in a rich (YT) medium. Grow a 2-3 ml overnight culture in 1X YT from a colony on a minimal plate, and use this to produce a log phage JM101/103 by 1/100 dilution in YT and subsequent growth. O/N cultures may be kept for up to a week at 4°C.

A drawback to the JM101/103 hosts has been observed in recent years. These hosts have active K12 restriction enzyme systems (EcoK) that restrict unmodified DNA. In addition, JM103 is lysogenic for phage P1 and hence has a second restriction system (EcoP1). If one attempts to transform unmodified DNA (grown in an $hsdM^-$ strain) into these hosts, a low transformation frequency can result. In most cases, however, these hosts have performed adequately.

New host strains, JM105, JM108, and JM109, which lack restriction systems, have been developed (11,12). These are not suppressor strains and therefore cannot be used to propagate mp7, mp8, or mp9. In practice, these new hosts do not seem to work as well, giving lower transformation efficiencies and plasmid yields. If one has particular trouble cloning a DNA fragment, it may be that it is rich in EcoK and EcoP1 sites. In this case it might be wise to do the cloning in JM105/109 and grow the phage (now modified) in JM101/103.

Contents

2.1 Isolation of DNA from Low Melting Point Agarose (LMA)

Low melting point agarose (from BRL) is a highly purified agarose that is non-toxic to many enzymes and has desirable melting properties (2). Even ligations done in 0.3% LMA suffer no decrease in efficiency. A LMA gel can be liquified by heating to 70°C for 5 min, and will remain molten as long as the temperature remains at or above 37°C. DNA contained in such a gel is not denatured by this heating, and can serve as an enzyme substrate after removal of ethidium bromide (EtBr) and addition of the appropriate 10X buffer. LMA has a higher cost and lower mechanical integrity than regular agarose. A 0.6% LMA is too soft to handle. Great care should be taken in removing the comb from a LMA gel to avoid tearing, although hardening the gel at 4°C helps to reduce this problem. The gel is run at room temperature exactly like a regular agarose gel.

Steps for fragment isolation from LMA (typical result is 60% yield or better):

1.) Run gel in 1X F-buffer without EtBr.

2.) Stain with 1 μg/ml EtBr, observe with long wave UV lamp.

3.) Cut out the band(s) to be isolated.

4.) 70°C, 5', Eppendorf tubes, 400 μl/tube.

5.) Remove from 70°C, quickly add 1 vol buffered phenol at 37°C, vortex (turns milky white); immediately microfuge 10', 4°C or room temp. (RT).

6.) Remove aqueous phase, avoiding fluffy white interface.

7.) Re-extract aqueous phase with 1 vol phenol, RT.

8.) Ether extract aqueous phase 2X.

9.) Add 1/10 vol 3 M NaAc, pH 6; 2.5 vol EtOH; mix well.

10.) Precipitate; dry ice, 15'-30', or -20°C overnight.

11.) Microfuge 10', 4°C; discard supernatant (SN).

12.) Wash 1X-2X with cold 75% EtOH. Dry pellet.

13.) Resuspend in desired volume of TE (10 mM Tris-HCl, pH 7.5, 1 mM EDTA).

14.) Quantitate by running an aliquot on a test gel next to a known amount of the digest from which the fragment was isolated.

2.2 Ligation Conditions

The following are ligation conditions that I have used to clone large fragments (up to 5.5 kb) into M13.

Reaction volume:	10 µl-20 µl
Vector DNA:	150 ng M13mp vector cut with the appropriate restriction enzyme(s). Gel purified linear fragment.
Insert DNA:	2X-3X molar excess over vector. Mass added depends on the size of the insert.
Buffer:	50 mM Tris-HCl, pH 7.5; 10 mM $MgCl_2$; 10 mM DTT; 0.5 mM ATP.
Ligase:	400 U (1 µl) NEB ligase or 2 U (1 µl) BRL ligase.

Ligate at room temperature 4 hrs to overnight. Do sticky end ligation whenever possible, since large fragments and blunt end ligation have not worked as well.

2.3 Transformation

1.) Dilute a JM101 O/N culture 1:100 in 1X YT. 5 ml/transformation.

2.) Incubate at 37°C with shaking to OD_{600} 0.2-0.4. About 1-1/2 hrs.

3.) Ice 10'.

4.) Pellet 7000 rpm, 5', 4°C, Sorvall swing out rotor or 2000 rpm, 6', 4°C in the J6.

5.) Discard SN. and resuspend in 1/2 volume of cold 50 mM $CaCl_2$. Be moderately gentle.

6.) Ice 20'.

7.) Pellet as before.

8.) Discard SN. Resuspend cells gently in 1/10 volume cold 50 mM $CaCl_2$. These are transformation competent cells; they can be kept at 4°C for at least several days with no loss in transformation efficiency, and can be stored indefinitely at -80°C after addition of glycerol. Transformation efficiencies are lower if the cells have been stored in glycerol, and I have always used fresh cells.

9.) Dispense 0.3 ml competent cells into tubes if plating is to be done on small (100 mm) plates or 0.5 ml if on large (150 mm) plates. On ice.

10.) Add ligation mix or 0.1 ng of M13 RF. The volume of the ligation mix should not exceed 1/2 the volume of the competent cells, and it is better not to exceed 1/4 volume. RF should give 1000 plaques/ng.

11.) Ice 40'.

12.) Heat shock 42°C, 2'. 45°C, 1-1/2 min is also effective.

13.) (Optional) If color reaction is desired, add:

10 µl 24 mg/ml IPTG (20 µl for large plates)

50 µl 2% Xgal in DMF (100 µl for large plates)

14.) Add heat shocked cells to 3 ml YT soft agar (8 ml for large plates) at 45°C. Mix and plate immediately.

15.) Incubate at 37°C, 6-12 hrs, upside down.

Uncut or re-ligated vector gives blue plaques with lac color indicator; recombinants give colorless plaques. Plaques may be screened by the Benton-Davis procedure like λ phage plaques (6).

IPTG = Isopropyl-beta-D-Thiogalactopyranoside (Sigma)

Xgal = 5-Bromo-4-Chloro-3-Indolyl-beta-D-Galactopyranoside (Sigma)

DMF = Dimethyl Formamide

2.4 Plaque Purification

Recombinants containing large inserts grow more slowly than M13 alone. Therefore, plaque purify the recombinants carefully before growing them up. Some apparent deletions were probably cases of improper plaque purification and subsequent overgrowth by contaminating vector. The methods described below have produced no deletions.

1.) Pick a colorless plaque with a sterile toothpick and drop into 1 ml TE (10 mM Tris-HCl, pH 7.5, 0.1 mM EDTA). You should get about 10^6 pfu.

2.) Serially dilute 10^{-2}, 10^{-3}, 10^{-4}.

3.) Plate 0.1 ml of the 10^{-3} and 10^{-4} dilutions with 200 µl JM101 log cells (OD_{600} 0.5). Sm. YT plates, 3 ml soft agar, IPTG, and Xgal.

4.) Incubate at 37°C, 6-12 hrs.

If the replating contains any blue plaques, as it usually does, repeat the above steps with one of the clear plaques from the replating until you have a plate containing only clear plaques. When all plaques are colorless, one can be used to start a large culture for an eventual RF prep.

2.5 Growth of Phage and Replicative Form (RF)

The following is the method for growing up M13 phage (7):

1.) Stab a plaque with a sterile toothpick.

2.) Drop toothpick into 2 ml early log phage JM101 in 1X YT. To prepare log cells, dilute an O/N culture 1/100, grow 1 hr. OD_{600} 0.05-0.2.

3.) Incubate at 37°C, 12-18 hrs with shaking.

4.) Transfer 1.5 ml to a microfuge tube. Microfuge 5'. Take SN.

5.) Titer SN. by making serial dilutions and plating on YT plates with 200 µl log cells. Expect about 10^{11} pfu/ml.

6.) Infect 50 ml early log phase JM101 (OD_{600} 0.2) with phage at a final concentration of 10^9 pfu/ml.

7.) Incubate at 37°C, 16-18 hrs with shaking.

8.) Spin out bacteria 7000 rpm, 5', 4°C Sorvall swing out rotor or 2000 rpm, 10', 4°C in the J6.

9.) Titer the supernatant. You should now have enough phage to do an RF prep.

RF preparation:

10.) For a large preparation, use 250 ml 1X YT per 2 liter flask.

11.) Add 2.5 ml JM101 O/N (1:100).

12.) Grow to an OD_{600} of 1.0. 2-3 hrs at 37°C with shaking.

13.) Infect with phage from step 9 at a final concentration of 10^{10} pfu/ml.

14.) Grow 2 hrs, 37°C, with shaking.

15.) Spin out cells, and proceed immediately with plasmid preparation.

2.6 Preparation of Cleared Lysate and Plasmid Preparation

1.) Begin with 250 or 500 ml culture after appropriate growth of plasmid or M13 RF.

2.) Pellet cells. 3500 rpm, 15', 4°C in 250 ml bottles, J6. Put water in the bottom of the buckets to cushion the bottles.

3.) Resuspend pellet in 50 ml TE, pH 8.5 (50 mM Tris, 1 mM EDTA).

4.) Pellet cells. 3500 rpm, 15', 4°C, J6.

5.) Resuspend pellet (500 ml or 250 ml pellet) in 16 ml 15% sucrose, 50 mM Tris-HCl, pH 8.5, 50 mM EDTA. Pipette up and down to resuspend. On ice.

6.) Add 4 ml freshly prepared 5 mg/ml lysozyme (Sigma) in the above buffer. Vortex gently and occasionally on ice, 10'.

7.) Add 12 ml 0.1% Triton X-100, 50 mM Tris-HCl, pH 8.5, 50 mM EDTA. Vortex gently. Incubate 10-15 min, 37°C to lyse cells. If proper lysis has occurred, the solution will become stringy and viscous from the released DNA.

8.) Load into an SW27 tube.

9.) Spin at 24,000 rpm, 45', 4°C in the ultracentrifuge.

10.) Decant the supernatant into a 50 ml Falcon centrifuge tube. This is the cleared lysate.

11.) Add, per ml of cleared lysate: 0.95 g CsCl
0.059 ml 2 mg/ml EtBr.

12.) Load into a VTi50 tube, discarding any excess material.

13.) Centrifuge at 45,000 rpm, 20 hrs, 20°C in the VTi50 rotor. Slow acceleration, no brake.

14.) Observe with a long wave UV lamp. Two bands should be visible. The upper band contains nicked plasmid, linear plasmid, and *E. coli* DNA. The lower band contains supercoiled, closed circular plasmid or M13 RF.

15.) Collect the lower band by dripping through an 18 g needle punctured through the side of the tube 5 mm below the band.

16.) Load band into a VTi80 tube and top off with a CsCl-EtBr solution prepared as follows: To 1 ml water add 1.059 g CsCl and 0.059 ml 2 mg/ml EtBr.

17.) Centrifuge at 60,000 rpm, 16 hrs, 20°C in the VTi80 rotor.

18.) Collect band with a syringe. The volume should be about 0.5 ml. There should be only one band this time.

19.) Butanol extract 5X-6X to get rid of EtBr. (n-Butanol-isopropanol, 1:1, saturated with a saturated TE-CsCl solution)

20.) Dialyze against TE (10 mM Tris-HCl, pH 7.5, 1 mM EDTA) to get rid of CsCl. Dialyze 10^6 or greater.

21.) Determine OD_{260} and OD_{280} to estimate the concentration and purity of the DNA. 260/280 should be 1.8 or greater. Yields are generally 500 µg to 1000 µg per liter of culture.

22.) Digest an aliquot with selected restriction enzymes, and check on an agarose gel.

2.7 Deletion Subcloning

Using the Sanger method, one can usually determine the nucleotide sequence of the first 200 bp of an insert cloned in M13. A few clones can be extended to about 300 bp, but beyond this the sequence is unreadable. In order to determine the sequence of a fragment of DNA longer than 300 bp, it is necessary to make cuts within the fragment and join the ends that are produced next to the primer binding region in an M13 phage. The 200 bp of sequence adjacent to the cut can then be read. If this is repeated for a number of different, appropriately spaced internal cuts, a series of overlapping 200 bp regions covering the full length of the fragment can be produced.

Deletion subcloning is a method for efficiently generating the overlapping subclones needed for sequence analysis of a large piece of DNA. This method replaces the "random shotgun" method previously used. The fragment to be sequenced is cloned into M13 and the double stranded RF is prepared by a plasmid prep. This parent subclone is used to produce a series of different sized subclones for sequencing, each of which has a deletion beginning at the Hind III site near the primer and terminating in the insert. The steps in this process are shown in Figure 2. The DNAase deletion subcloning method is similar to several published procedures (8-10).

A. DNAase Digestion

In a manganese buffer, DNAase I makes mostly double stranded breaks instead of the single stranded nicks that it produces in a magnesium buffer (8). The enzyme apparently cuts randomly, although there may be preferred sites. Conditions are chosen in this step to produce about one double stranded DNAase cut per phage molecule. Only those molecules that have the DNAase cut in the insert will produce viable phages later on, because a DNAase cut in the vector disrupts necessary phage genes.

1X DNAase buffer: 20 mM Tris-HCl, pH 7.6, 3 mM $MnCl_2$

1X Reconstitution buffer: 1X DNAase buffer + 500 µg/ml BSA

Make 5X DNAase buffer fresh each day, since it turns brown on storage, even when frozen. DNAase I (Sigma) is made as an inactive 1 mg/ml stock in 0.01 M HCl. Divide into 50 µl aliquots and store frozen at -80°C. To reactivate, add 0.95 ml reconstitution buffer to a 50 µl aliquot and incubate on ice 3-6 hrs. Just before use, dilute a portion of this to 0.5 ng/µl (1:100 dilution) in reconstitution buffer.

Digest 30 µg of parent subclone RF with 3 ng DNAase I (100 pg DNAase/µg DNA in a volume of about 200 µl. The buffer is 1X DNAase. Incubate 15 min at 25°C (room temp.). Stop the reaction by adding 0.5 M EDTA, pH 7.5 to 10 mM.

Titrate the amount of DNAase needed to give the maximum conversion to the linear form the first time you use a stock of DNAase. For my DNAase stock, 100 pg/µg DNA gives satisfactory results, with little variation from clone to clone.

B. Frequent Cutting Endonuclease Digestion

A panel of frequent cutting (4 base pair recognition sequence) restriction endonucleases can be used instead of DNAase I to make random cuts in the parent subclone. The following panel of enzymes has been used successfully: Alu I, Hae III, Rsa I, Dde I, Hinf I, and Sau3A I.

1X Restriction Nuclease Buffer: 10 mM Tris-HCl, pH 7.5
10 mM $MgCl_2$
50 mM NaCl
1 mM DTT

Titrate each enzyme separately. For Alu I, 1 hr unit per microgram of DNA incubated for 30 min at 37°C gave maximum conversion to the linear form. Most of the other enzymes required 0.1-0.6 U per µg of DNA. Digest 6 µg of DNA with each of the six enzymes (36 µg total) under the appropriate conditions for 30 min at 37°C, add EDTA to 10 mM to stop the reaction, and pool the six digestions.

C. Gel Electrophoresis

Load into two wells of an 8-well, horizontal, 0.7% LMA gel. Run in 1X F-buffer without EtBr. Include the linear and supercoiled parent clone as markers. Electrophorese O/N at 40 V. Stain the gel with 1 µg/ml EtBr and observe under a long wave length UV lamp. The gel should look like Figure 3. Isolate the linear band from the DNAase lanes by the method in Section 2.1. Resuspend the DNA in 50 µl TE (10 mM Tris-HCl, pH 7.5, 1 mM EDTA). Take 2 µl, and run this alongside known amounts of linear (Hind III digested) parent clone to estimate the yield. Use 100 ng, 200 ng, and 400 ng as your markers. The yield should be 5-10 µg.

D. Hind III Digestion

Any restriction enzyme which cleaves the parent clone only once between the primer and the inserted DNA fragment may be used instead of Hind III. For example, if you cloned a Bam HI fragment containing an internal Hind III site but no Sal I site, you would use Sal I (and in fact, could not use Hind III).

Digest to completion by incubating the DNA (5-10 µg) with 200 U (hr units)

of Hind III in a 100 µl of 1X RN (see 2.7B) for 1 hr at 37°C. Add 0.5 M EDTA to 10 mM to stop the reaction. Phenol-Sevag extract, EtOH ppt, wash pellet 1X with 75% EtOH, dry, resuspend in 20 µl TE. Run a minigel to confirm that the Hind III cut to completion.

E. Klenow Fill-in of Ends

The large fragment of *E. coli* DNA polymerase (Klenow fragment) is used to make all the DNA fragments blunt ended in preparation for later blunt end ligation.

Reaction conditions:
50 mM NaCl
10 mM Tris-HCl, pH 7.6
10 mM $MgCl_2$
1 mM DTT
1 mM each dATP, dCTP, dGTP, TTP

Incubate with 10 U Klenow (Boehringer) 30 min, 30°C. Add 0.5 M EDTA to 10 mM to stop the reaction.

F. Gel Electrophoresis

Load 4 µg DNA into a single well of a 12-well, horizontal, 0.6% LMA gel. The running buffer is 1X F-buffer without EtBr. Include linear parent clone and linear M13mp8 as markers. Electrophorese 20 hrs, 50 V.

Cut out the section of the gel containing the markers, stain with 1 µg/ml EtBr in 1X F-buffer, and destain in 1X F-buffer. Be very careful handling the gel; use a thin piece of plastic as a scoop to pick up the gel, do not attempt to move it with just your hands. Return the gel section next to the unstained DNAase/Hind III lane. If the entire gel were stained, it would look like Figure 4. Cut 1-2 mm gel slices in the DNAase/Hind III lane using the markers as a guide. Start 1 mm above the linear parent clone, and end 4 mm below the linear mp8. Cut slices with a gel slicer, or by hand with a scalpel. A parent clone with a 3 kb insert should give about 15-20 slices.

G. Treatment of Gel Slices

Place each slice in a separate Eppendorf tube, and add water to a volume of 250 µl. Melt the gel by heating to 70°C for 5 min. Transfer to 37°C. Add 10X ligation buffer. Add ATP to 0.5 mM, and water to 300 µl. Remove from water bath and add 0.5-1.0 µl T4 DNA ligase (NEB 400 U/µl). Ligate at room temperature 4 hrs - O/N.

Transformation: Heat ligation mix to 70°C to melt any reformed matrix. Cool to 37°C, then add 100-150 µl of this to 0.5 ml competent cells (see Section 2.3) and plate on large YT plates in 8 ml of YT soft agar. Incubate overnight, 37°C. There should be 30-200 plaques per plate. Start by doing T-ladders on four plaques from each plate to get a general idea of overlaps and to identify nearly identical clones.

H. Note: Keep in mind that all the sequencing subclones generated from a parent clone will give you sequence on the same strand. In order to sequence the complementary strand, you will need to clone the fragment in both orientations. Given the difficulties in reading dideoxy sequencing gels, it is necessary to have the complementary strand to be 100% sure of your sequence.

2.8 Growth of Phage and Preparation of DNA for Sequencing

1.) Stab a sterile toothpick into a plaque.

2.) Drop into 2 ml 1:100 dilution of JM101 O/N in 1X YT.

3.) Incubate at 37°C, 6-10 hrs, with shaking.

4.) Remove 1.5 ml into Eppendorf tube.

5.) Microfuge 10', room temperature.

6.) Transfer 1.2 ml to a new tube (no bacteria!).

7.) Add 300 µl of: 20% PEG 8000

2.5 M NaCl

8.) Mix by inversion. Inc. 15'-30' at room temperature.

9.) Microfuge 10', room temperature.

10.) Remove S.N. Spin again 1', remove all of SN. with a pipetman or a drawn out Pasteur pipette.

11.) Suspend phage pellet in 160 µl TES:

20 mM Tris-HCl, pH 7.5

10 mM NaCl

0.1 mM EDTA.

(You may save a few µl for re-infection)

12.) Add 100 µl buffered phenol.

13.) Vortex until mixed, let stand 5', vortex again.

14.) Microfuge 2'-4', room temperature or 4°C.

15.) Remove 140 µl SN. (no phenol or interface).

16.) (Optional) Transfer to a new tube, and add 70 µl Sevag (Chloroform:isoamyl alcohol 24:1). Vortex. Microfuge 2'. Remove 120 µl SN. (no $CHCl_3$ or interface).

17.) Add 1/10 vol 3 M NaAc, pH 6, 2.5 vol EtOH; mix well.

18.) Precipitate: dry ice 15'-30', or -20°C overnight.

19.) Microfuge 10', 4°C, discard SN.

20.) Wash 1X-2X with cold 75% EtOH. Dry pellet.

21.) Resuspend in 25 µl TE (10 mM Tris-HCl, pH 7.5, 0.1 mM EDTA).

22.) (Optional) Add 175 µl TE and repeat steps 17-20.

23.) Store at 4°C or -20°C.

Note: PEG in the phage miniprep will greatly inhibit the sequencing reactions, and will produce a smeary, unreadable gel. Be sure to get all of the supernatant of the PEG ppt.

2.9 Sequencing Reactions

-ddNTP's from Collaborative Research or PL-Biochemicals

-Primer (15-mer) from Collaborative Research

-Klenow from Boehringer

-Use ^{32}P-CTP or ^{32}P-ATP to label DNA. If you use ATP, change the N° mixes and chase with ATP instead of CTP.

Make 10 mM stocks of the four dNTP's the the four ddNTP's in TE (10 mM Tris-HCl, pH 7.5, 1 mM EDTA), and divide into 50 µl aliquots. Store at -80°C, except for the aliquot you are using, which can be stored at -20°C. Make primer as a 10 ng/µl stock in water. Store at -20°C.

Make up the following solutions. All are in water unless otherwise indicated.

10X Hin buffer: 500 mM NaCl
70 mM Tris-HCl, pH 7.5
100 mM $MgCl_2$
30 mM DTT

R35 0.5 mM dATP (5 µl 10 mM stock diluted 20X)
R36 0.5 mM dCTP (")
R37 0.5 mM dGTP (")
R38 0.5 mM dTTP (")

R39 Å
1 µl R35 (A)
20 µl R37 (G)
20 µl R38 (T)
20 µl 10X Hin

R40 Č
20 µl R35 (A)
20 µl R37 (G)
20 µl R38 (T)
20 µl 10X Hin

R41 $\mathring{G}$	20 µl R35 (A)
	1 µl R37 (G)
	20 µl R38 (T)
	20 µl 10X Hin
R42 $\mathring{T}$	20 µl R35 (A)
	20 µl R37 (G)
	1 µl R38 (T)
	20 µl 10X Hin

R43	0.38 mM ddATP	dilute 10 mM 1:30
R44	0.25 mM ddCTP	dilute 10 mM 1:40
R45	0.25 mM ddGTP	dilute 10 mM 1:40
R46	1.0 mM ddTTP	dilute 10 mM 1:10

The optimal ratios of ddNTPs should be determined for each stock of ddNTP by doing sequencing reactions on M13 at several ddNTP dilutions (1:3, 1:10, 1:30, 1:90). The dilutions for R43-R46 are those determined for my stocks of ddNTPs.

Next, make the following four N-mixes. 2 µl per reaction.

A' = 1 µl R39 + 1 µl R43

C' = 1 µl R40 + 1 µl R44

G' = 1 µl R41 + 1 µl R45

T' = 1 µl R42 + 1 µl R46

Keep R35-R46 on ice, and store at -20°C. These solutions should be good for about two weeks. Make A', C', G', T' fresh the day you use them.

1.) Label tubes, and add 1 µl template DNA (from 2.8) and 2 µl of the appropriate N-mix. Room temperature.

2.) Make up Premix 1, 2 µl per reaction. Keep on ice until ready to add to reaction. Add Klenow last. Receipe for 80 µl:

50 µl water

10 µl 10X Hin

13 µl ^{32}P-CTP (10 mCi/ml, 400 Ci/mMol. aqueous, Amersham)

4 µl Primer (10 ng/µl

3 µl Klenow (5 U/µl, Boehringer)

3.) Add 2 µl Premix 1 to each reaction.

4.) Incubate 30°C, 30 min.

5.) Chase with 1 µl R36 per reaction.

6.) Incubate 30°C, 15 min.

7.) Add 12 µl Formamide-dye loading buffer to stop reaction.

8.) Denature 90°-95°C for 3 min; quick chill to 0°C.

9.) Load 2-4 µl per lane on a sequencing gel.

2.10 Sequencing Gels

Read the first 60-80 bp of an insert on a 40 cm 8% acrylamide sequencing gel. When the xylene cyanol has moved 23 cm the filled-in Hind III site will be at the bottom of the gel. To read the next 80-100 bp, load the same sample on an 80 cm 5% acrylamide sequencing gel. The xylene cyanol is run 60-65 cm to get the proper overlap. If it is necessary to read further, load another 80 cm 5% gel and allow the xylene cyanol to run 100 cm. If further sequence is needed, another loading may be done and the xylene cyanol run 140 cm.

A typical 40 cm gel is 40 cm x 30 cm x 0.04 cm. The following is the gel recipe for one gel. This recipe includes enough extra to avoid running out while pouring the gel.

50 g Urea (Ultrapure Schwartz-Mann)

7.6 g Acrylamide (Biorad or equivalent)

0.4 g Bis-acrylamide (Biorad)

10 ml 10X TEB

water to 100 mls

filter through Whatman 1 mm

0.7 ml 10% Ammonium Persulfate

10 µl TEMED

Pour gel between appropriately sealed, siliconized glass plates. Insert comb, and add extra TEMED at the teeth. Allow to polymerize for at least one hour.

The recipe for a typical 80 cm 5% sequencing gel is as follows:

100 g Urea

9.5 g Acrylamide

0.5 g Bis-acrylamide

13.4 ml 10X TEB

water to 200 ml

filter (Whatman)

1.5 ml 10% Ammonium Persulfate

40 µl TEMED

The buffer concentration is different for long and short gels. For 40 cm gels, use 90 mM Tris, and for 80 cm gels, use 60 mM Tris. Run short gels at 1200-1400 V, and long gels at 2000-2500 V.

2.11 Appendix

10X F-buffer:
400 mM Tris base
200 mM Sodium acetate
10 mM EDTA
titrated to pH 7.2 with glacial acetic acid

10X TEB:
0.9 M Tris base
1.0 M Boric Acid
10 mM EDTA
should be pH 8.3

Agarose Gel Electrophoresis Loading Buffer:
45% Sucrose
1X F-buffer
0.03% Bromphenol Blue
Add 1/2-1/5 vol to DNA before loading

YT Medium (1X):
8 g/l Bacto tryptone
5 g/l Bacto yeast extract
5 g/l NaCl

Minimal Medium:
10.5 g/l K_2HPO_4
4.5 g/l KH_2PO_4
1.0 g/l $(NH_4)_2SO_4$
0.5 g/l Na Citrate$\cdot 2H_2O$
0.2 g/l $MgSO_4 \cdot 7H_2O$*
5.0 µg/l Thiamine-HCl*
2.0 g/l Glucose*

*Sterilized separately as a concentrated stock

Agar plates: 15 g/l Bacto-agar in appropriate medium (YT or minimal)

YT soft agar: 8 g/l Bacto-agar in YT medium

Formamide/dye Sequencing Gel Loading Buffer:

90% Deionized Formamide (BRL ultrapure)

10 mM EDTA

10 mM NaOH

0.1-0.3% each, Xylene Cyanol and Bromphenol Blue

Buffered Phenol: Ultrapure or redistilled phenol equilibrated with 10 mM Tris, pH 8, 1 mM EDTA. Store frozen in 10 ml aliquots.

2.12 References

1. 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.

2. Weislander, L. 1979. A simple method to recover intact high molecular weight RNA and DNA after electrophoretic separation in low gelling temperature agarose gels. *Anal. Biochem.* **98**:305.

3. Wartell, R. M. and W. S. Reznikoff. 1980. Cloning DNA restriction endonuclease fragments with protruding, single-stranded ends. *Gene* **9**:307.

4. Maniatis, T., E. F. Fritsch and J. Sambrook. 1982. *Molecular Cloning, A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp. 93-94, 113-114, 250-251.

5. Mandel, M. and A. Higa. 1970. Calcium dependent bacteriophage DNA infection. *J. Mol. Biol.* **53**:154.

6. Benton, W. D. and R. W. Davis. 1977. Screening λgt recombinant clones by hybridization to single plaques *in situ*. *Science* **106**:180.

7. Hines, J. C. and D. S. Ray. 1980. Construction and characterization of new coliphage M13 cloning vectors. *Gene* **11**:207.

8. Fischauf, A. M., H. Garoff and H. Lehrach. 1980. A subcloning strategy for DNA sequence analysis. *Nucleic Acids Res.* **8**:5541.

9. Hong, G. F. 1982. A systematic DNA sequencing strategy. *J. Mol. Biol.* **158**:539.

10. Barnes, W. B., M. Bevan and P. H. Son. 1983. Kilo-sequencing: Creation of an ordered nest of asymmetric deletions across a target sequence carried on phage M13. *Meth. Enzymol.* **101**:98.

11. Norrander, J., T. Kempe and J. Messing. 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene* **26**:101.

12. Yanisch-Perron, C., J. Vieira and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**:103.

Figure 1. DNA sequence of the polylinker region of three even-numbered M13mp-series vectors showing the primer binding site and the direction of synthesis for chain termination sequencing. The corresponding odd-numbered vectors, mp9, mp11, and mp19, respectively, contain the polylinker in the opposite orientation.

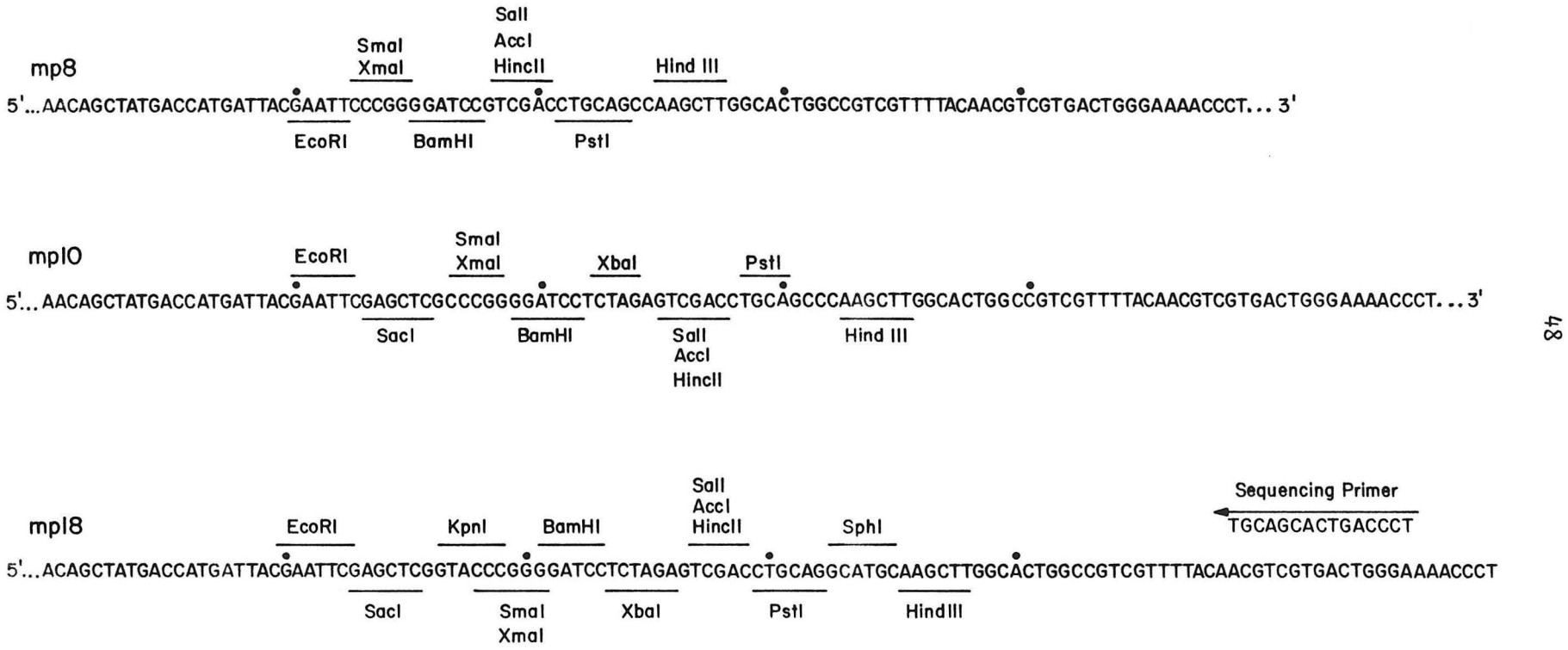
mp8
SalI
AccI
HincII
SmaI
XmaI
Hind III
5'...AACAGCTATGACCATGATTACGAATTCCCGGGGATCCGTCGACCTGCAGCCAAGCTTGGCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCT... 3'
EcoRI
BamHI
PstI
mp10
EcoRI
SmaI
XmaI
XbaI
PstI
5'... AACAGCTATGACCATGATTACGAATTCGAGCTCGCCCGGGGATCCTCTAGAGTCGACCTGCAGCCCAAGCTTGGCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCT...3'
SacI
BamHI
SalI
AccI
HincII
Hind III
mp18
SalI
AccI
HincII
Sequencing Primer
TGCAGCACTGACCCT
EcoRI
KpnI
BamHI
SphI
5'...ACAGCTATGACCATGATTACGAATTCGAGCTCGGTACCCGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCT
SacI
SmaI
XmaI
XbaI
PstI
HindIII

Figure 1

Figure 2. Deletion subcloning. (A) A parent subclone containing an Eco RI (RI)-Hind III (H) insert (open box) too large to be sequenced completely from the universal primer. Hatched area within the insert is internal sequences that will be brought next to the primer by the deletion subcloning method. (B) Parent subclone after partial digestion with DNAase I or frequent cutting restriction endonucleases. This represents just one of many molecules in a heterogeneous mix to be size selected later, as the other molecules will have single random cuts at other places in the insert and phage genome. This is the linear form isolated in Figure 3. (C) After digesting with Hind III and making the ends of each DNA molecule blunt ended by filling in with dNTPs and Klenow, the mixture is ready for size selection as in Figure 4. (D) After appropriate size selection, the molecule is recircularized, bringing the internal (hatched) sequences next to the primer binding site for sequence determination.

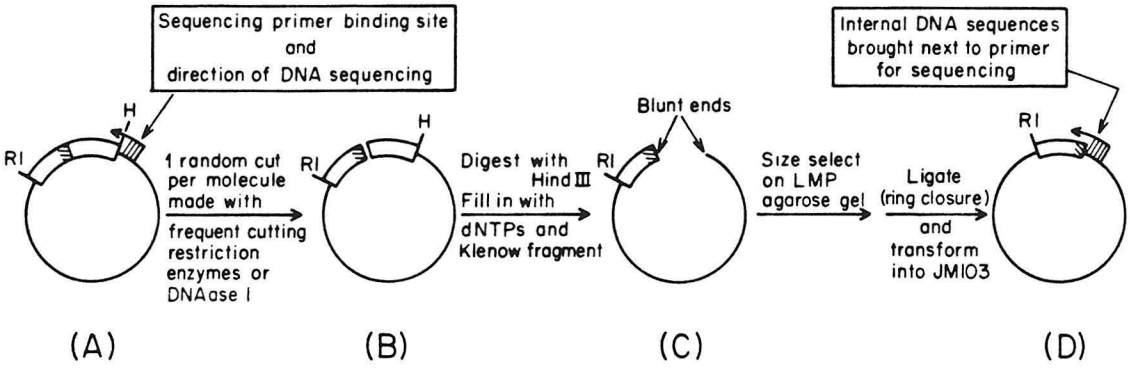
Sequencing primer binding site
and
direction of DNA sequencing
H
RI
1 random cut
per molecule
made with
frequent cutting
restriction
enzymes or
DNAase I
Digest with
Hind III
Fill in with
dNTPs and
Klenow fragment
Blunt ends
Size select
on LMP
agarose gel
Ligate
(ring closure)
and
transform
into JM103
Internal DNA sequences
brought next to primer
for sequencing
(A)
(B)
(C)
(D)

Figure 2

Figure 3. Partial digestion with DNAase I or frequent cutting restriction endonucleases. (A) The supercoiled parent RF. (B) The result of proper partial digestion. Some supercoil remains, but most of the DNA is converted to the linear form with some nicked circles inevitable. Insufficient digestion would contain mostly supercoil and overdigestion would result in no supercoil and a smear of fragments smaller than the linear form.

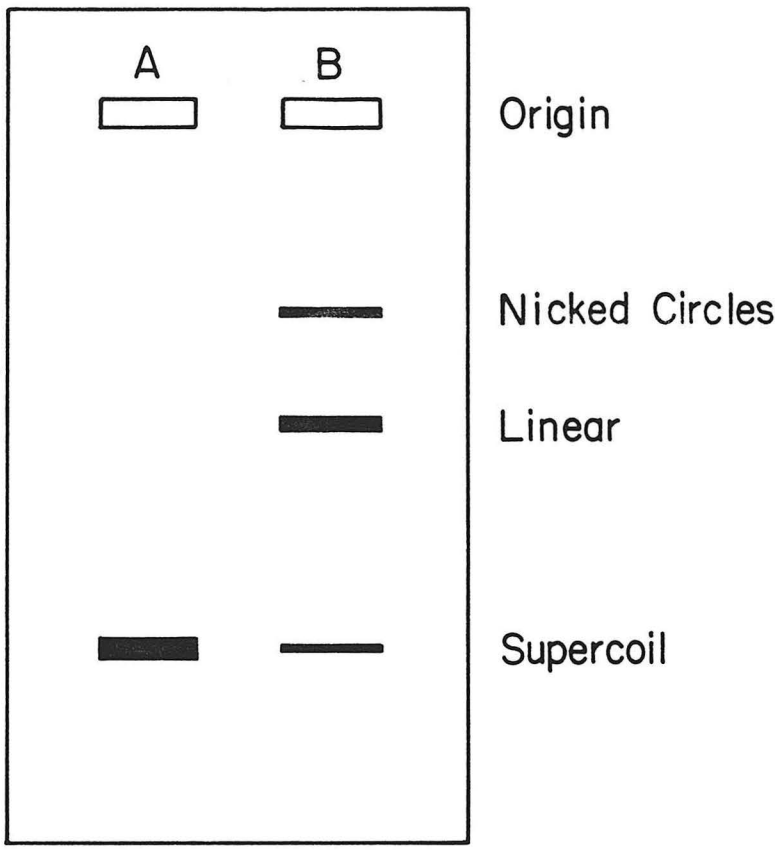

Figure 3

Figure 4. Size selection on LMA. (A) Linear vector without insert. (B) Linear parent subclone. (C) Parent subclone after partial digestion and Hind III digestion. Lane C is cut with a scalpel or gel slicer as indicated, and each gel slice treated separately as in 2.7-G.

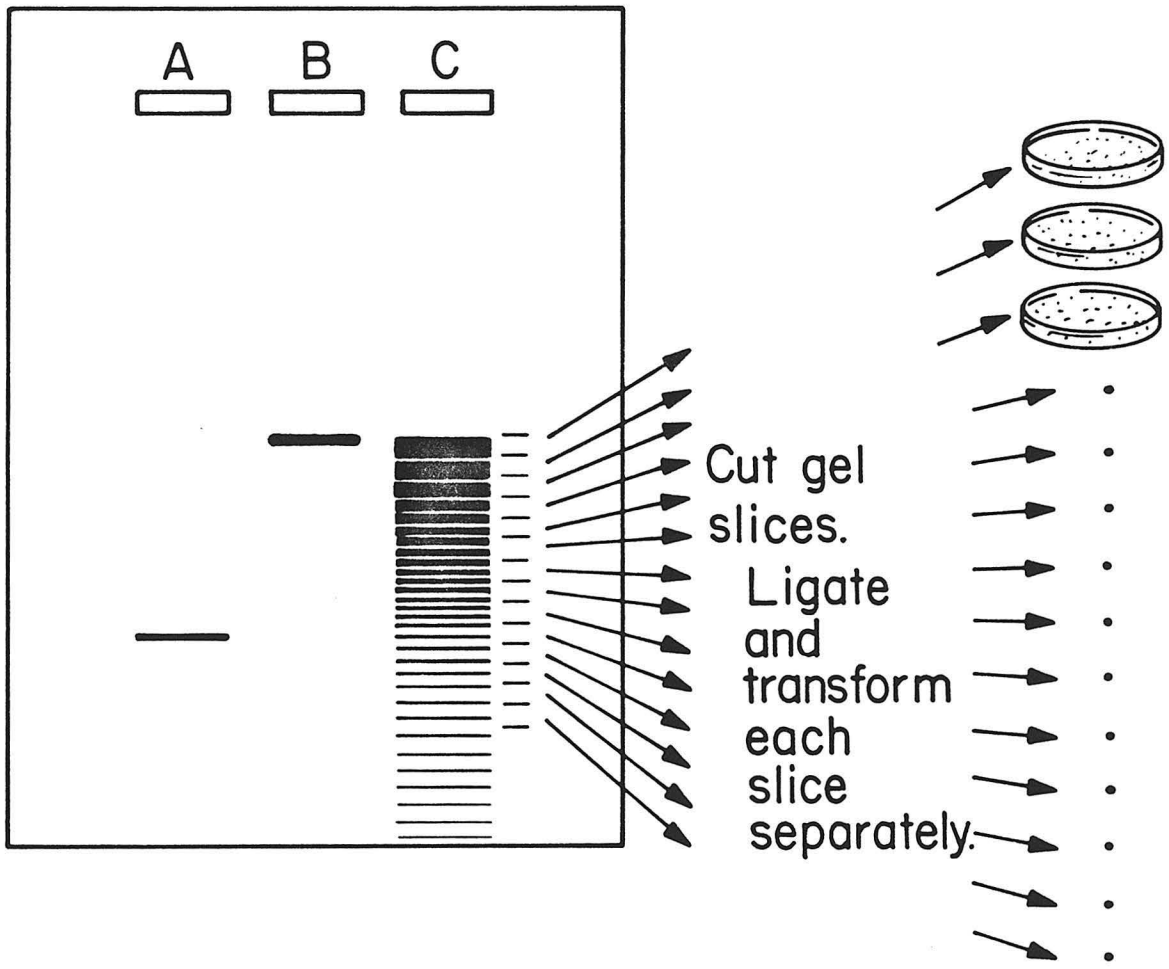
A
B
C
Cut gel
slices.
Ligate
and
transform
each
slice
separately.

Figure 4

Chapter 3

STRUCTURE OF A GENE ENCODING A MURINE THYMUS LEUKEMIA ANTIGEN AND ORGANIZATION OF *Tla* GENES IN THE BALB/c MOUSE*

By DOUGLAS A. FISHER, STEPHEN W. HUNT III, and LEROY HOOD

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

Running title: *Tla* GENE SEQUENCE AND ORGANIZATION

*Supported by grants from the National Institutes of Health; S.W.H. is a Leukemia Society postdoctoral fellow.

J. Exp. Med., **162**, 528-545

Introduction

The murine class I molecules are a family of cell surface glycoproteins that includes the transplantation antigens H-2K, H-2D, and H-2L, as well as the lymphoid differentiation antigens Qa-1, Qa-2, Qa-3, and thymus leukemia (TL) antigen. These antigens are encoded in the major histocompatibility complex (MHC) on chromosome 17 (Fig. 1). Class I molecules have a molecular weight of 40,000-45,000 daltons, and associate non-covalently with β_2-microglobulin. The transplantation antigens are important cell-cell recognition molecules for cytotoxic T lymphocytes (1). The function of the Qa/TL antigens is unknown, but their presence on lymphoid cells, coupled with their homology to transplantation antigens, suggests that they too may be involved in cell-cell interactions in the immune system.

TL antigen is a class I molecule of special interest because of its varied patterns of expression in both normal and leukemic cells. Transplantation antigens are expressed on virtually all somatic cells, while TL antigens are expressed only on thymocytes, some thymic leukemias, and activated T lymphocytes (2-4). In prothymocytes, TL antigen expression is induced in response to thymic hormones during the maturation of these cells into thymocytes (5, 6). Mature T cells that migrate to the peripheral lymphoid system no longer express detectable TL antigen, except when stimulated to proliferate (4). Six serologically defined alleles of *Tla* exist, and mice having the three most commonly studied alleles, Tla^a, Tla^b, and Tla^c, differ in the quantity of TL antigen expressed on their thymocytes as well as in serological determinants (7, 8). Thymocytes from Tla^a mice express about 20 times as much TL antigen as those from Tla^c mice, while thymocytes from Tla^b mice are generally considered to be TL^-; however, one report suggests that Tla^b thymocytes may express very low amounts of TL antigen (8). In contrast, leukemias of all three strains may express TL antigen at levels comparable with Tla^a thymocytes.

In the past four years, our understanding of class I genes has increased dramatically due to the isolation of cDNA clones from both human and mouse class I genes (9-11). Steinmetz et al. (12, 13) used a class I cDNA clone to isolate 36 class I genes from λ-phage and cosmid genomic libraries constructed from BALB/c mouse DNA. The chromosomal location of each gene was mapped using strain specific restriction enzyme polymorphisms and recombinant mouse strains, and five of these genes mapped to the *H-2* complex, while the remainder mapped to the *Qa-2,3* and *Tla* regions (14). To identify the genes encoding serologically defined products, each of the cloned genes was transformed into mouse fibroblasts (L cells) by DNA-mediated gene transfer and the resulting transformants were examined with monoclonal antibodies against H-2K^d, H-2L^d, H-2D^d, Qa-2,3, and TL antigen (15). The K^d, L^d, D^d, *Qa-2,3* and two *Tla* genes were identified by this method. Since then, numerous class I genes have been sequenced including the three BALB/c transplantation antigen genes and two *Qa-2,3* region genes (12, 16-20). However, none of the serologically defined *Qa/Tla* genes have been characterized to date.

In this report, we present the complete genomic nucleotide sequence and analysis of a BALB/c (Tla^c) gene encoding a TL antigen. The polypeptide chain is encoded in six exons homologous to the first six exons of other class I genes. Southern blot analyses using *Tla*-specific probes subcloned from this gene have allowed us to reorganize the *Tla* region gene clusters described earlier (13), and have reduced the number of class I genes detected in the BALB/c mouse to 33: 18 in the *Tla* region, 10 in the *Qa-2,3* region, and five in the *H-2* region.

Materials and Methods

Materials. Restriction endonucleases and T4 DNA ligase were obtained from New England Biolabs. The large fragment of E. coli DNA polymerase was from Bethesda Research Laboratories. Deoxynucleoside triphosphates (dNTPs) were

obtained from Sigma, dideoxynucleoside triphosphates and the 15 nucleotide sequencing primer from Collaborative Research, and α-^{32}P dNTPs from Amersham. E. coli strain JM103, phage M13mp8, and phage M13mp10 were from Bethesda Research Laboratories.

Methods. Commonly used recombinant DNA procedures such as gel electrophoresis, Southern blotting, nick translation, growth of plasmids and phage, and restriction endonuclease digestions followed previously described protocols (21, 22).

DNA Sequence Analysis. The sequencing reactions for the chain termination or chemical degradation methods were carried out according to previously published procedures (23, 24). Overlapping subclones for chain termination sequencing in phage M13 were generated using a deletion subcloning method similar to several published methods (25-27).

Results and Discussion

Gene 17.3A Encodes a Serologically Defined TL Antigen. Clone 17.3 was previously shown to transform mouse L cells to the TL^+ phenotype as assayed by monoclonal anti-TL antibodies in a radioimmune assay (15). A partial restriction map of the eukaryotic insert of clone 17.3 is shown in Fig. 2. Since clone 17.3 contains two class I genes (denoted 17.3A and 17.3B), it was necessary to determine which of the genes was responsible for the TL-transforming activity. Digestion of clone 17.3 with the restriction endonucleases SacII or BamHI, which cut within 17.3B, did not affect the TL-transforming activity, whereas digestion with EcoRI, which cleaves only within 17.3A, abolished the TL-transforming activity (R. Goodenow, unpublished results). Also shown in Fig. 2 is pTLA.1, a pBR322 subclone derived from clone 17.3 that contains only gene 17.3A. Subclone pTLA.1 contains the TL-transforming activity of clone 17.3, and this activity is abolished by digestion with EcoRI. Since EcoRI digests the insert DNA only

once—within gene 17.3A—we conclude that gene 17.3A encodes a serologically defined TL antigen.

The 17.3A Polypeptide Chain is Encoded by Six Exons Homologous to the First Six Exons of Transplantation Antigen Genes. The nucleotide sequence of gene 17.3A was determined using the sequencing strategy shown in Fig. 2. The DNA sequence is shown in Fig. 3. The first five exons of gene 17.3A correspond both in sequence and in the position of intron-exon borders with the first five exons of other class I genes. These exons encode a hydrophobic leader peptide, three external protein domains of about 90 amino acids each, and a hydrophobic transmembrane segment. A cytoplasmic protein domain is encoded entirely by exon 6, unlike transplantation antigens where the cytoplasmic domain is usually encoded by three exons.

The first exon of 17.3A encodes a hydrophobic signal peptide that is either 24 or 26 amino acids long, depending on which of two in frame AUG start codons initiates translation. Ordinarily, the first AUG encountered in a eukaryotic messenger RNA initiates translation (31). The reason for the uncertainty is that a potential TATA box (bases 196-199) is located so close to the beginning of exon 1 that the first AUG (bases 220-222) might not be included in the mRNA, assuming, as with other eukaryotic genes, transcription begins 26-34 bp downstream of the TATA box (32). There are other AT-rich regions that might serve as TATA boxes (bases 144-149, 154-160), and although a consensus sequence CCAAT box is not found, several similar sequences are associated with the potential TATA boxes (CCATT at bases 93-97, CAAGT at bases 174-178, CAAA at bases 181-184).

Exons 2, 3, and 4 encode three protein regions of about 90 amino acids each, denoted $\alpha 1$, $\alpha 2$, and $\alpha 3$, respectively (33). These three regions comprise all of the TL molecule that is exposed on the cell surface. The positions of cysteine residues are homologous to those of other class I genes: both the $\alpha 2$ and $\alpha 3$

regions contain cysteine residues about 60 amino acids apart, suggesting that each contains an intrachain disulfide linkage, while the α1 region contains no cysteines. The TL molecule encoded by gene 17.3A contains two potential carbohydrate attachment sites, (Asn-X-$\overset{Ser}{Thr}$), both in the α1 domain, at amino acids 86 and 90. It is not known from biochemical data how many carbohydrate moieties are present on TL molecules from BALB/c (Tla^c) mice, but there is only one moiety on TL molecules encoded by the Tla^a allele (34).

Exon 5 encodes the transmembrane segment of the TL polypeptide. The translated amino acid sequence contains 38 amino acids, including a span of 20 uncharged and generally hydrophobic amino acids (15 hydrophobic, three glycine, and two serine residues) that presumably span the lipid bilayer and terminate in a cytoplasmic region with two positively charged amino acids.

Three regions homologous to the sixth, seventh, and eighth exons of other class I genes can be identified in gene 17.3A, and are indicated in Fig. 3. However, in gene 17.3A, an in-frame termination codon is reached in exon 6, so this exon encodes the entire cytoplasmic domain and the carboxy terminus of the TL polypeptide chain. This fact is corroborated by the DNA sequence of a closely related Tla^b cDNA clone isolated using a transmembrane probe from gene 17.3A. In this cDNA clone, the protein coding sequence terminates in exon 6, and no RNA splice sequences are used downstream of exon 6 (Obata, Y., Chen, Y., Stockert, E. and Old, L., personal communication). The cytoplasmic domain of gene 17.3A is two amino acids shorter than the cytoplasmic domain of the L^d molecule, and has only 30% (7/23) amino acid homology with the L^d molecule.

The low sequence homology between transplantation antigens and gene 17.3A in the cytoplasmic region may be a clue to the different behavior of TL molecules and transplantation antigens in the cell membrane. In a process known as antigenic modulation, TL antigens are rapidly lost from the cell surface after

incubation with anti-TL antibody (even monovalent antibody), while transplantation antigens are not (35-37). The presence of a very different cytoplasmic domain suggests that TL antigens may interact with the cytoskeleton or other cytoplasmic components differently than H-2 antigens, but further studies, such as exon shuffling experiments, will be needed to localize the structural component responsible for the antigenic modulation effect.

Comparison of the Nucleotide and Amino Acid Sequence of 17.3A with K^d, D^d, L^d, and a Qa-2,3 Gene. The DNA sequence of gene 17.3A was compared with sequences from the following class I genes: K^d (17), L^d (16) D^d (20), the *Qa-2,3* gene Q6 (27.1) (12), K^b (18), and the Q10 gene encoding a secreted class I molecule from the *Qa-2,3* region (19, 38). None of these genes is significantly more closely related to gene 17.3A than to any other, and all are more closely related to each other than any one is to gene 17.3A. With an overall homology to the K^d gene of only 74% at the DNA level, gene 17.3A is the most divergent murine class I gene sequenced to date. Table I shows the DNA and amino acid homologies of gene 17.3A with four BALB/c class I genes: K^d from the K end of the *H-2* locus, L^d and D^d from the D end of *H-2*, and gene Q6, from the *Qa-2,3* region.

All four genes are approximately equally homologous to gene 17.3A, but each exon is not equally conserved (Table I). The fourth exon is the most highly conserved exon, probably due to selective pressure on the sequence of the $\alpha 3$ region encoded by exon 4, since the $\alpha 3$ region binds non-covalently with β_2-microglobulin (39). Exons 1 and 5 encoding the hydrophobic leader and transmembrane peptides are only moderately conserved, since the sequence constraints on these two regions probably require only hydrophobic properties. The cytoplasmic exon is the least conserved and, as mentioned above, this may reflect different interactions with cytoskeletal elements. The degree of homology of exons 2 and 3 with the genes in Table I is intermediate between that seen for

exon 4 and exon 5. In H-2 molecules, the α1 and α2 regions are the site of recognition by cytotoxic T lymphocytes (40), but since the function of the α1 and α2 regions in TL antigens is unknown, it is difficult to estimate the nature of the selective pressure operating on these regions.

The homologies noted above are consistent with earlier peptide mapping analyses. Peptide mapping studies showed that out of 13-19 tryptic peptides resolved on an HPLC cation exchange column, three peptides co-eluted between the H-2D^d molecule and a *Tla*^{a} gene product, for an estimated peptide homology of 20%. In identical experiments, different H-2 antigens were found to have about 40% peptide map homology, and allelic forms of TL were found to share 70-80% of their peptides (41, 42). The translated amino acid sequences of the *D*^{d} and 17.3A genes share three predicted tryptic peptides. Two of the peptides come from the α3 region and one is the N-terminal peptide. The actual percent peptide homology between the D^d and TL antigens is somewhat lower than the previous estimates because the HPLC system used to separate the tryptic peptides could only resolve 16-19 peptides, while 30-43 peptides are predicted from the translated amino acid sequence.

The three *H-2* genes are more closely related to each other at the DNA level (>90% homology) than the *H-2* genes are to 17.3A (74%). In addition, the *Qa-2,3* region genes are much more homologous to the *H-2* genes (89%) than to 17.3A (74%). These data probably mean that *Tla* and *H-2* genes diverged from a common ancestor earlier than H-2 genes diverged from one another. The greater sequence homology of the *Qa-2,3* and *H-2* genes suggests that *H-2* and *Qa-2,3* genes diverged more recently than the *H-2* and *Tla* genes.

The Third Intron of Gene 17.3A Contains Large DNA Deletions and/or Insertions. The third intron is related in an interesting manner to the third intron of the *K*^{d} gene (Fig. 4A). At a point 95 bp from the 5' end of this intron, gene

17.3A contains 1,093 bp of sequence not found in the K^d gene. Whether this is an insertion in gene 17.3A or a deletion in the K^d gene is unknown, but the 1,093 bp segment contains a 65 bp thymidine-rich (76%) sequence at its 5' end and an Alu-like repeat element near its 3' end. The segment is flanked by inexact 10 bp inverted repeats which are also present in the K^d gene (Fig. 4B). The third intron of the K^d gene is about the same length as the 17.3A third intron because the K^d intron contains 1,130 bp segment of DNA not found in 17.3A (Fig. 4A). This DNA segment is bordered by inexact 16 bp direct repeats (Fig. 4C), has a simple sequence (the repeated dinucleotide TC) at its 5' end, and has an Alu repeat near its 5' end. The location of this insertion/deletion is 30 bp 5' to the fourth exon. Because of the presence of inverted and direct repeats bordering the 1,093 bp and 1,130 bp regions, it is tempting to speculate that these regions were derived from transposon-like insertions.

Part of Gene 17.3A Appears to be the Product of a Gene Conversion Event. The 17.3A introns are only about 60% homologous to those of other class I genes, while the introns of the three BALB/c transplantation antigen genes are about 90% homologous to each other (20). The one exception to the lower sequence homology between the noncoding sequences of gene 17.3A and the *H-2* genes is the first 71 bp of the fourth intron (Fig. 5A), located immediately adjacent to the highly conserved fourth exon. It is not surprising that the fourth exon is highly conserved, because this exon encodes the $\alpha 3$ region that binds with β_2-microglobulin, but it is very surprising to find the adjacent intron so highly conserved (97% with the L^d gene) since the other introns are only about 60% homologous.

To test whether the fourth exon of gene 17.3A is more highly conserved than would be expected from natural selection on the protein sequence, we counted the number of silent site mutations (base substitutions that do not alter the protein

sequence) between the L^d and 17.3A genes. Silent site mutations accumulate with time at a constant rate, and are believed to be independent of selection at the protein level, so by using their frequency one can estimate the time elapsed since two related genes shared a common ancestor (43). There are mutations in 43% of the silent sites in exons 2 and 3 of gene 17.3A but only in 11% of the silent sites of exon 4 when compared with the L^d gene (Fig. 5B). This disparity (43% vs. 11%) is very large, and corresponds to a statistically significant (>3 σ) difference in divergence times estimated for these different regions by the calculation of Kimura (43). This means that the fourth exon of gene 17.3A shared a common ancestor with the L^d gene much more recently than the second and third exons. The last finding, coupled with the observation that the first 71 bp of the fourth intron are highly conserved, leads us to the hypothesis that a region including all of exon 4 and part of the fourth intron was transferred to the 17.3A gene as a gene conversion event from one of the *H-2* genes. A gene conversion of this size (~350 bp) would serve to decrease the sequence polymorphism of the fourth exon, in contrast to the role envisioned for very small (<50 bp) gene conversions in exons 2 and 3 of mutant *H-2* genes which, it has been suggested, would increase the polymorphism of these regions (18, 44, 45).

Fourth exon gene conversions or gene corrections may have occurred in other class I genes. All the murine class I genes examined have a low percentage of silent site mutations in exon 4 (6% average) compared with the percentage in exons 2 and 3 (14% average). When the K^d, L^d, and Q6 genes are compared with two functional human transplantation antigen genes (46, 47), no significant disparity is seen in the number of silent site mutations in exons 2 and 3 (29.4% average) and exon 4 (27.8% average). This observation is also consistent with a gene correction or conversion model because the human and mouse genes have been separated by speciation and could not have engaged in gene conversion.

Gene conversions of variable size occur between class I genes, but it is unclear how many of these events are biologically significant. The small gene conversions mentioned above may contribute to the polymorphism of transplantation antigens. However, it is difficult to imagine why natural selection would not be sufficient to maintain the protein sequence of the α3 domain. Gene conversion may be a result rather than a cause of the high DNA sequence homology of exon 4. Since gene conversions presumably require an initial base pairing between homologous sequences, conversions may occur more frequently in exon 4 because of the high sequence homology of this exon with other class I genes.

Hybridization With Tla-specific Probes Detects a Small Subset of Class I Genes That Are Polymorphic Between Tla Alleles. In order to study the number and diversity of genes related to 17.3A, as well as to create DNA probes suitable for examining RNA species expressed in TL^+ cells, two subclones, pTLA.4 and pTLA.5, were constructed from the coding regions of gene 17.3A that exhibited minimal homology with other class I genes (Fig. 2). The 5' probe, pTLA.4 (bases 577-892), contains parts of the second and third exons, as well as the entire second intron. This probe is about 70% homologous with the K^d or L^d genes in the coding sequences. The 3' probe, pTLA.5 (bases 3155-3795), contains the transmembrane exon and is only about 50% homologous to the corresponding regions in the K^d and L^d genes. Both probes hybridize to BALB/c thymus poly(A)$^+$ RNA, but not to BALB/c liver or spleen RNA. The probes also do not hybridize to thymus, liver, or spleen RNA from C57BL/6, a TL^- strain (data not shown).

Subclones pTLA.4 and pTLA.5 were hybridized to genomic DNA from four strains of mice representing the *Tla* alleles Tla^a, Tla^b, Tla^c, and Tla^d. As can be seen in Fig. 6, there is size polymorphism in the hybridizing restriction fragments. This is surprising since others have had relative difficulty finding polymorphic

restriction fragments in the *Tla* region compared with the *H-2* region (14). Four *Tla*-specific bands are found in the Southern analysis of Tla^c mouse DNA. These data are summarized in Table II. One band corresponds to gene 17.3A, the gene from which the probes were subcloned, and another corresponds to gene 24.8, the other *Tla* gene identified by transformation into L cells (15). However, gene 24.8 by itself could not encode a TL antigen because DNA sequence analysis indicates that this gene is a pseudogene due to the presence of numerous in-frame stop codons (D. Fisher, unpublished results). Gene 17.3A hybridizes with both the 5' and 3' *Tla* probes while gene 24.8 hybridizes only to the 3' probe. The other two hybridizing genes are of two types: one is like 17.3A because it hybridizes to both pTLA.4 and pTLA.5, while the other is like 24.8-like and hybridizes only to pTLA.5.

New Linkage of Tla Clusters. When the two *Tla* probes were hybridized to a panel of cosmid clones containing all 36 class I genes of the gene clusters described by Steinmetz et al. (13), seven genes, rather than the expected four, were detected. This observation prompted us to examine these cluster linkages, and led to the discovery of three cloning artifacts, on cosmids 1.1, 8.3, and 20.1. Each of these cosmids contains a large piece of DNA with a restriction map identical to that of cosmids in other gene clusters and a shorter piece of DNA with a unique restriction map. These cosmids probably result from the cloning of two non-contiguous pieces of eukaryotic DNA ligated together during the construction of the cosmid library. We isolated low copy number probes from near the border between the putative non-contiguous DNA fragments, and by Southern blotting analysis showed that restriction fragments of the size predicted by cosmids 1.1, 8.3, and 20.1 were not present in the genome, while fragments from overlapping cosmids (in agreement with the cluster map in Fig. 7) were present in the genome (data not shown). The three cosmids in question have been previously noted as possible cloning artifacts (14, 48).

Elimination of the artifactual constructs allows several new overlaps between the gene clusters described by Steinmetz et al. (Fig. 7). Clusters 3, 7, 8 and 4 can now be joined into a single cluster 160 kb in length, which we denote cluster A. Cluster A contains the three artifactual constructs. A second cluster of 77 kb, denoted cluster B, is formed from clusters 12 and 5. This linkage was previously overlooked because the 10 kb overlapping region contains only three mapped restriction sites. A third Tla cluster, cluster 10, a single cosmid clone containing an incomplete class I gene, is unchanged from the earlier study. This new linkage is consistent with genome blots using both *Tla*-specific probes because four genes (T1, T3, T11, and T13) hybridize to pTLA.5 and two genes (T3 and T13) hybridize to pTLA.4 (Table II, Figs. 6 and 7). The sizes of the restriction fragments hybridizing to these probes as determined in genomic blots (Table II) are consistent with the sizes predicted from the restriction maps of the corresponding cosmid clones. Similar cluster linkage results have recently been independently obtained by others (49). The new gene linkage reduces the number of class I genes found in the BALB/c mouse to 33: 18 in the *Tla* region, 10 in the *Qa-2,3* region, and five in the *H-2* region.

Examination of the restriction map of the gene clusters reveals large (20-40 kb) stretches of DNA that appear to be related. These regions are indicated as bars above the genes in Fig. 7. The relatedness of some of these duplicated regions has been established not only by restriction map similarity, but also by hybridization with low copy DNA probes. For example, the four genes hybridizing with the 3' *Tla* probe, pTLA.5, are contained in two 40 kb segments (hatched bars in Fig. 7A and 7B) that appear to have duplicated at some time in the past. Although it is not possible to prove the mechanism by which these regions arose, duplication by homologous but unequal crossover seems to be a likely explanation.

Tla^b Mice Contain Fewer Tla Genes Than Tla^c Mice, and Gene 17.3A Is One of Those Absent in Tla^b. Recently, Weiss et al. (48) have isolated cosmid clones containing class I genes from the C57BL mouse (Tla^b), and have linked the *Tla* region class I genes into one gene cluster. Twenty-six class I genes were found in C57BL as compared to 33 in BALB/c. Most of the additional BALB/c genes were found in the *Tla* region. The Tla^b gene cluster can be aligned with cluster A, and all the genes in cluster A are found in the Tla^b cluster (Fig. 8). However, no genes corresponding to cluster B or cluster 10 are found among the C57BL cosmid clones. Genomic Southern blots (Fig. 6) using the *Tla*-specific probes confirm that at least two genes (genes T11 and T13 in Fig. 7) in BALB/c are not found in Tla^b, and both of these genes are located in cluster B. Although these two genes are the only ones confirmed to be absent in Tla^b, it seems probable that none of the genes in cluster B or cluster 10 are present in C57BL mice.

Interestingly, gene 17.3A (gene T13 in Fig. 7) is absent in Tla^b. Could this be the explanation for the TL^- phenotype of Tla^b thymocytes? Clearly, since Tla^b leukemias can express TL antigen, Tla^b mice must have a structural gene encoding a TL antigen. Since allelic forms of TL antigen share 70-80% of their tryptic peptides (41) and are therefore very closely related, we expect that both the *Tla* specific coding region probes would hybridize to the gene encoding the leukemic Tla^b gene product. If this is the case, then the T3 gene of the *Tla* gene cluster described by Weiss et al. (48) (homologous to gene T3 in Fig. 7) encodes the Tla^b leukemic TL molecule, since this is the only C57BL gene to hybridize with pTLA.4. It will be interesting to determine whether the same gene that is active in Tla^c thymocytes is induced in Tla^c TL^+ leukemias. Perhaps the Tla^b gene and its homolog in Tla^c are expressed in leukemic cells, while 17.3A (absent in Tla^b) is expressed in thymocytes.

Summary

We have determined the DNA sequence of a gene encoding a TL antigen in the BALB/c mouse, and have more definitively mapped the cloned BALB/c *Tla* region class I gene clusters. Analysis of the sequence shows that the *Tla* gene is less closely related to the *H-2* genes than *H-2* genes are to one another or to *Qa-2,3* region genes. The *Tla* gene, 17.3A, contains an apparent gene conversion. Comparison of the BALB/c *Tla* genes with those from C57BL shows that BALB/c has more *Tla* region class I genes and that one of the genes absent in C57BL is gene 17.3A.

Acknowledgements

We thank Karyl Minard and Debbie Maloney for expert technical assistance, and Connie Katz for preparation of the manuscript. The authors are grateful to Jerry Siu and Drs. E. Rothenberg, M. Kronenberg, I. Stroynowski and N. Costlow for helpful comments on the manuscript.

Note added in proof: A detailed restriction map of a cDNA clone isolated from a BALB/c thymocyte cDNA library using pTLA.5 as a probe (Shen, F.-W., personal communication) confirms the 3' splicing pattern proposed in this paper. As in the Tla^b cDNA clone (see text), none of the potential splice sequences (indicated by arrows in Fig. 3) downstream of exon 6 are used.

References

1. 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.

2. 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. Nat. Cancer Inst.* **31**:977.

3. Flaherty, L. 1980. The Tla region antigens. *In* Role of the Major Histocompatibility Complex in Immunology. M.E. Dorf, editor. Garland Press, New York, pp. 33-57.

4. Cook, R. G., and N. F. Landolfi. 1983. Expression of the thymus leukemia antigen by activated peripheral T lymphocytes. *J. Exp. Med.* **158**:1012.

5. Scheid, M. P., G. Goldstein, and E. A. Boyse. 1978. The generation and regulation of lymphocyte populations: evidence from differentiative induction systems in vitro. *J. Exp. Med.* **147**:1727.

6. Lepault, F., and I. Weissman. 1981. An in vitro assay for thymus-homing bone marrow cells. *Nature (Lond.).* **293**:151.

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

8. Michaelson, J., E. A. Boyse, M. Chorney, L. Flaherty, E. Fleissner, U. Hämmerling, C. Reinisch, R. Rosenson, and F.-W. Shen. 1983. The biochemical genetics of the Qa-Tla region. *Transplantation Proc.* **15**:2033.

9. Ploegh, H. L., H. T. Orr, and J. L. Strominger. 1980. Molecular cloning of a human histocompatibility antigen cDNA fragment. *Proc. Nat. Acad. Sci. USA.* **77**:6081.

10. Sood, A. K., D. Pereira, and S. M. Weissman. 1981. Isolation and partial nucleotide sequence of a cDNA clone for human histocompatibility antigen HLA-B by use of an oligodeoxynucleotide primer. *Proc. Nat. Acad. Sci. USA.* **78**:616.

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

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

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

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

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

16. 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.

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

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

19. Mellor, A. L., E. H. Weiss, M. Kress, G. Jay, and R. A. Flavell. 1984. A nonpolymorphic class I gene in the major histocompatibility complex. *Cell.* **36**:13.

20. Sher, B. T., R. Nairn, J. Colligan, and L. Hood. 1985. DNA sequence of the mouse D^d gene. *Proc. Nat. Acad. Sci. USA.* **82**:1175.

21. Davis, R. W., D. Botstein, and J. R. Roth. 1980. Advanced Bacterial Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

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

23. 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.

24. Maxam, A. M. and W. Gilbert. 1980. Sequencing end labeled DNA with base specific chemical cleavages. *Meth. in Enzymol.* **65**:499.

25. Frischauf, A. M., H. Garoff, and H. Lehrach. 1980. A subcloning strategy for DNA sequence analysis. *Nucleic Acids Res.* **8**:5541.

26. Hong, G. F. 1982. A systematic DNA sequencing strategy. *J. Mol. Biol.* **158**:539.

27. Barnes, W. B., M. Bevan, and P. H. Son. 1983. Kilo-sequencing: Creation of an ordered nest of asymmetric deletions across a target sequence carried on phage M13. *Methods in Enzymol.* **101**:98.

28. Davis, M. M., K. Calame, P. W. Early, D. L. Livant, R. Joho, I. L. Weissman, and L. Hood. 1980. An immunoglobulin heavy chain gene is formed by at least two recombinational events. *Nature (Lond.).* **283**:733.

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

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

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

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

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

34. Rothenberg, E., and D. Triglia. 1981. Structure and expression of glycoproteins controlled by the Qa-1^a allele. *Immunogenetics* **14**:455.

35. Boyse, E. A., L. J. Old, and S. Luell. 1963. Antigenic properties of experimental leukemias. II. Immunological studies in vivo with C57BL/6 radiation-induced leukemias. *J. Nat. Cancer Inst.* **31**:987.

36. Boyse, E. A., E. Stockert, and L. J. Old. 1967. Modification of the antigenic structure of the cell membrane by thymus leukemia (TL) antibody. *Proc. Natl. Acad. Sci. USA.* **58**:954.

37. Stackpole, C. W., J. B. Jacobson, and M. P. Lardis. 1974. Antigenic modulation in vitro. I. Fate of thymus leukemia (TL) antigen-antibody complexes following modulation of TL antigenicity from the surfaces of mouse leukemia cells and thymocytes. *J. Exp. Med.* **140**:939.

38. Kress, M., D. Cosman, G. Khoury, and G. Jay, 1983. Secretion of a transplantation related antigen. *Cell.* **34**:189.

39. Yokoyama, K., and S. G. Nathenson. 1983. Intramolecular organization of class I H-2 MHC antigens; localization of the alloantigenic determinants and the β_2-m binding site to different regions of the H-2 K^b glycoprotein. *J. Immunol.* **130**:1419.

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

41. Yokoyama, K., E. Stockert, L. J. Old, and S. G. Nathenson. 1981. Structural comparisons of TL antigens derived from normal and leukemia cells of TL^+ and TL^- strains and relationship to genetically linked H-2 major histocompatibility complex products. *Proc. Nat. Acad. Sci. USA.* **78**:7078.

42. McIntyre, K. R., U. Hämmerling, J. W. Uhr, and E. S. Vitetta. 1982. Structural analysis of thymus leukemia (TL) antigens in the mouse. *J. Immunol.* **128**:1712.

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

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

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

46. Sodoyer, R., M. Damotte, T. L. Delovitch, J. Trucy, B. R. Jordan, and T. Strachan. 1984. Complete nucleotide sequence of a gene encoding a functional human class I histocompatibility antigen (HLA-CW3). *EMBO J.* **3**:879.

47. Strachan, T., R. Sodoyer, M. Damotte, and B. R. Jordan. 1984. Complete nucleotide sequence of a functional class I HLA gene, HLA-A3: implications for the evolution of HLA genes. *EMBO J.* **3**:887.

48. 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.

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

Table I

DNA and Amino Acid Homologies: 17.3A vs. 4 BALB/c Class I Molecules

	K^d		L^d		D^d		27.1 (Qa-2,3)	
	DNA	Protein	DNA	Protein	DNA	Protein	DNA	Protein
Exon 1 (leader)	64%	52%	53%	43%	64%	54%	56%	29%
Exon 2 (α1)	71%	56%	70%	56%	72%	58%	70%	56%
Exon 3 (α2)	74%	62%	75%	65%	74%	65%	75%	63%
Exon 4 (α3)	91%	83%	92%	85%	92%	89%	91%	86%
Exon 5 (TM)	64%	35%	64%	38%	65%	41%	62%	35%
Cytoplasmic	42%	35%	42%	30%	43%	30%	45%	31%
TOTAL	74%	61%	74%	61%	75%	64%	74%	60%

Table II

BALB/c Class I Genes Hybridizing to pTLA.4 and pTLA.5

Gene Name	Gene in Figure 7	pTLA.4	pTLA.5	Size of hybridizing fragment		
				Hind III	Bam HI	Kpn I
17.3A-like	T3	+	+	13 kb	15.5 kb	7.8 kb
17.3A	T13	+	+	8 kb	15.5 kb	7.8 kb
24.8	T1		+	8.3 kb	~20 kb	3.7 kb
24.8-like	T11		+	8.3 kb	12 kb	3.7 kb

Figure 1. Genetic map of the MHC. Genes indicated with open boxes encode class I molecules. The order of loci within brackets is not known. Distances are in centiMorgans (cM).

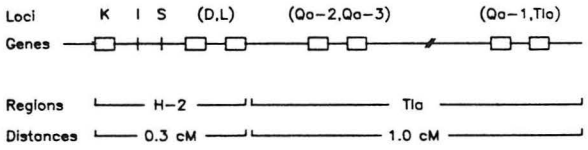
Loci
K
I
S
(D,L)
(Qa-2,Qa-3)
(Qa-1,Tla)
Genes
Regions
H-2
Tla
Distances
0.3 cM
1.0 cM

Figure 1

Figure 2. Restriction map and sequencing strategy for gene 17.3A. (A) Partial restriction map of the insert of Charon 4A clone 17.3A. Open boxes indicate regions hybridizing with class I cDNA clones (11), and the arrows indicate the transcriptional orientation of the two class I genes, 17.3A and 17.3B. EcoRI sites marked with an asterisk result from the addition of EcoRI linkers during the construction of the phage library (28). (B) Partial restriction map of the eukaryotic insert of pTLA.1, a subclone derived from 17.3. pTLA.1 contains a 16.5 kb ClaI-BamHI fragment that includes 3.5 kb of sequence from the left arm of Charon 4A (not shown) subcloned into ClaI-BamHI digested pBR322. (C) Restriction map and sequencing strategy for gene 17.3A showing the locations of TL antigen coding sequences (solid boxes) and Alu-like repetitive elements (hatched boxes) related to either the mouse B1 or B2 sequences (29, 30). Sequencing done by the chain termination method (23) is indicated with straight arrows, while that determined by the chemical degradation method (24) is indicated by wavy arrows. The parts of the gene that were subcloned to make the two *Tla*-specific probes, pTLA.4 and pTLA.5, are shown as brackets above the map. The scale is in kilobases.

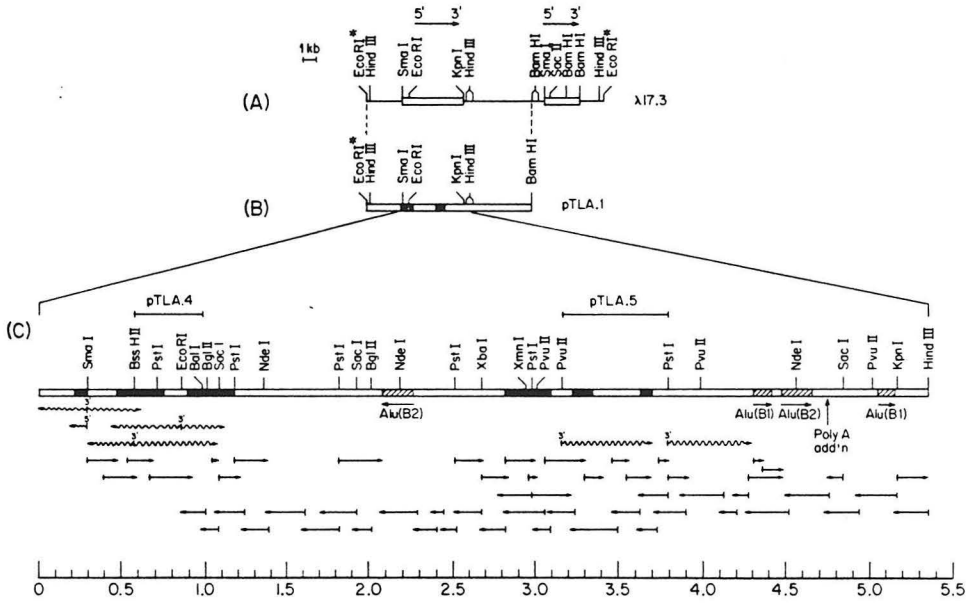
1kb
(A)
λ17.3
(B)
pTLA.1
(C)
pTLA.4
pTLA.5
Alu(B2)
Alu(B1)
Alu(B2)
Poly A add'n
Alu(B1)
0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5

Figure 2

Figure 3. Sequence of gene 17.3A. Coding regions are shown as triplets with the translated amino acids (single letter code) above the DNA sequence. In the promoter region, the putative TATA box is highlighted by a box and several potential CCAAT box sequences are underlined. Near the 3' end of the gene, arrows indicate probable splice junctions terminating exon 6, bordering exon 7, and beginning exon 8. However, it is not known if these downstream splices are used at all. In the Tla^b cDNA clone, none of the splice junctions indicated by arrows are used (see text). The untranslated portion of exon 6, as well as exon 7 are underlined, and the poly(A) addition signal is indicated by a box. Two poly(A) addition signals occur before this point in the 3' untranslated region, but these sequences are located in an Alu repeat and probably do not initiate poly(A) addition. At least two class I genes lacking the Alu repeat (K^d, Q10) utilize a poly(A) signal very close to the one indicated by a box in 17.3A. Ambiguous base code: N = A, C, G, T; P = A, G; Y = C, T; S = G, C; R = A, T.

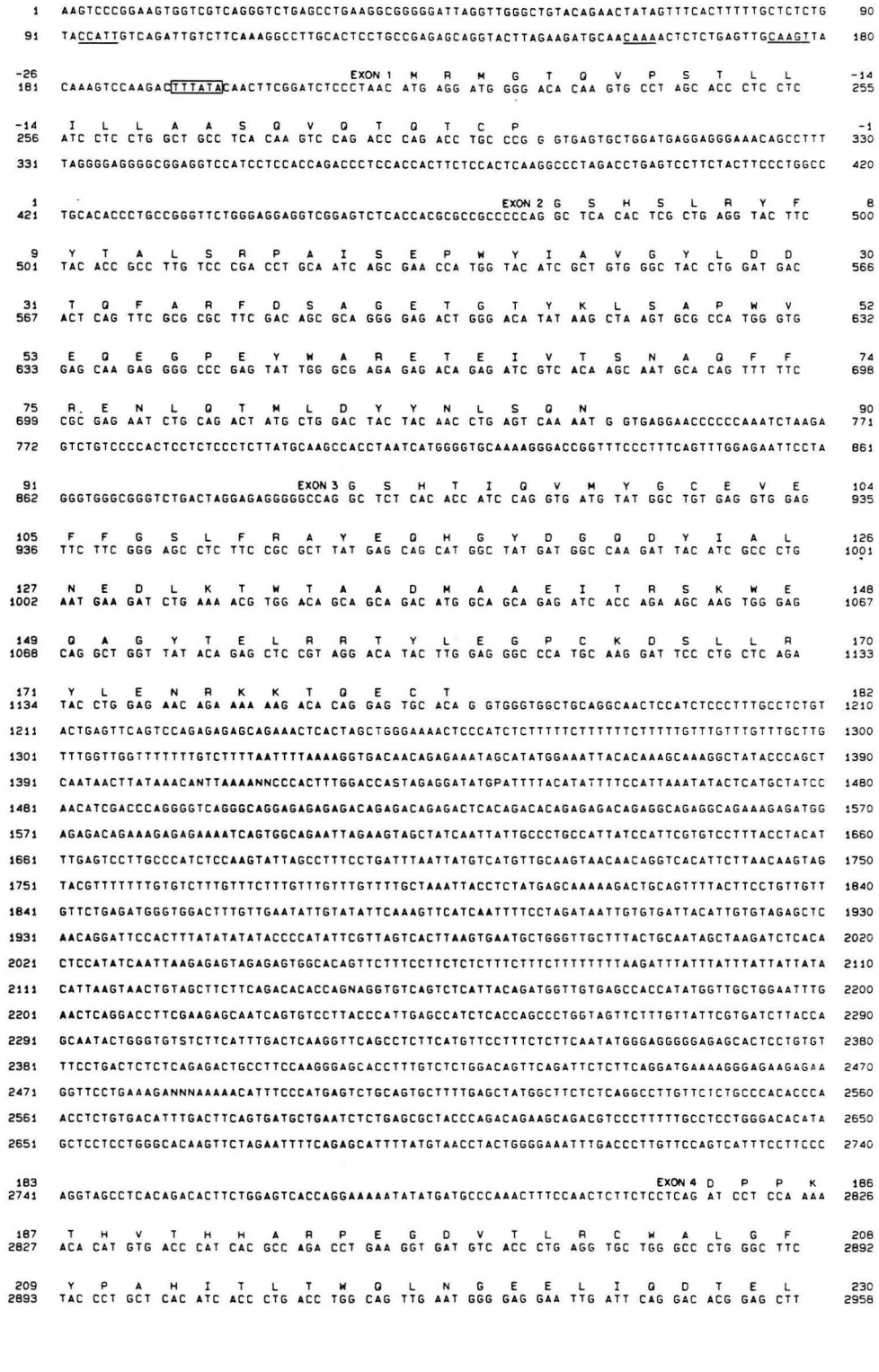

Figure 3, pg. 1

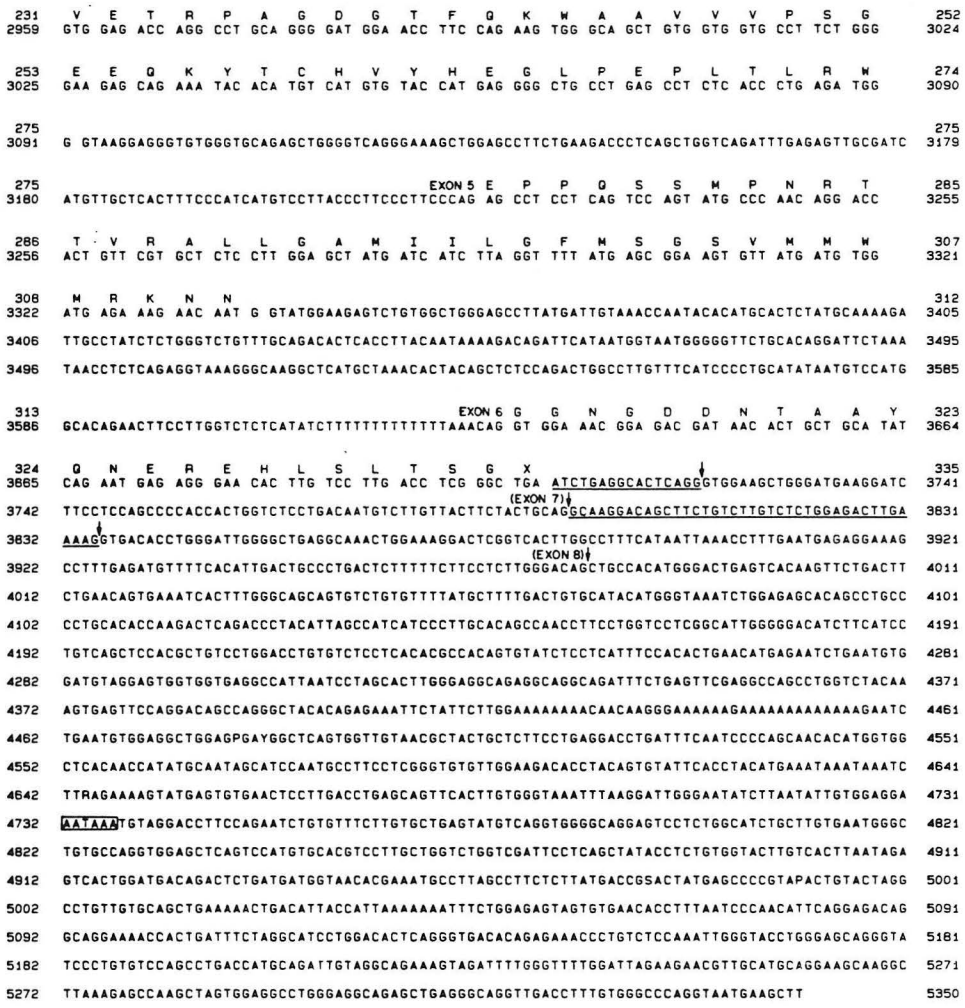
231 V E T R P A G D G T F Q K W A A V V V P S G 252
2959 GTG GAG ACC AGG CCT GCA GGG GAT GGA ACC TTC CAG AAG TGG GCA GCT GTG GTG GTG CCT TCT GGG 3024
253 E E Q K Y T C H V Y H E G L P E P L T L R W 274
3025 GAA GAG CAG AAA TAC ACA TGT CAT GTG TAC CAT GAG GGG CTG CCT GAG CCT CTC ACC CTG AGA TGG 3090
275 275
3091 G GTAAGGAGGGTGTGGGTGCAGAGCTGGGGTCAGGGAAAGCTGGAGCCTTCTGAAGACCCTCAGCTGGTCAGATTTGAGAGTTGCGATC 3179
275 EXON 5 E P P Q S S M P N R T 285
3180 ATGTTGCTCACTTTCCCATCATGTCCTTACCCTTCCCTTCCCAG AG CCT CCT CAG TCC AGT ATG CCC AAC AGG ACC 3255
286 T V R A L L G A M I I L G F M S G S V M M W 307
3256 ACT GTT CGT GCT CTC CTT GGA GCT ATG ATC ATC TTA GGT TTT ATG AGC GGA AGT GTT ATG ATG TGG 3321
308 M R K N N 312
3322 ATG AGA AAG AAC AAT G GTATGGAAGAGTCTGTGGCTGGGAGCCTTATGATTGTAAACCAATACACATGCACTCTATGCAAAAGA 3405
3406 TTGCCTATCTCTGGGTCTGTTTGCAGACACTCACCTTACAATAAAAGACAGATTCATAATGGTAATGGGGGTTCTGCACAGGATTCTAAA 3495
3496 TAACCTCTCAGAGGTAAAGGGCAAGGCTCATGCTAAACACTACAGCTCTCCAGACTGGCCTTGTTTCATCCCCTGCATATAATGTCCATG 3585
313 EXON 6 G G N G D D N T A A Y 323
3586 GCACAGAACTTCCTTGGTCTCTCATATCTTTTTTTTTTTTTAAACAG GT GGA AAC GGA GAC GAT AAC ACT GCT GCA TAT 3664
324 Q N E R E H L S L T S G X 335
3665 CAG AAT GAG AGG GAA CAC TTG TCC TTG ACC TCG GGC TGA ATCTGAGGCACTCAGGGTGGAAGCTGGGATGAAGGATC 3741
(EXON 7)
3742 TTCCTCCAGCCCCACCACTGGTCTCCTGACAATGTCTTGTTACTTCTACTGCAGGCAAGGACAGCTTCTGTCTTGTCTCTGGAGACTTGA 3831
3832 AAAGGTGACACCTGGGATTGGGGCTGAGGCAAACTGGAAAGGACTCGGTCACTTGGCCTTTCATAATTAAACCTTTGAATGAGAGGAAAG 3921
(EXON 8)
3922 CCTTTGAGATGTTTTCACATTGACTGCCCTGACTCTTTTTCTTCCTCTTGGGACAGCTGCCACATGGGACTGAGTCACAAGTTCTGACTT 4011
4012 CTGAACAGTGAAATCACTTTGGGCAGCAGTGTCTGTGTTTTATGCTTTTGACTGTGCATACATGGGTAAATCTGGAGAGCACAGCCTGCC 4101
4102 CCTGCACACCAAGACTCAGACCCTACATTAGCCATCATCCCTTGCACAGCCAACCTTCCTGGTCCTCGGCATTGGGGGACATCTTCATCC 4191
4192 TGTCAGCTCCACGCTGTCCTGGACCTGTGTCTCCTCACACGCCACAGTGTATCTCCTCATTTCCACACTGAACATGAGAATCTGAATGTG 4281
4282 GATGTAGGAGTGGTGGTGAGGCCATTAATCCTAGCACTTGGGAGGCAGAGGCAGGCAGATTTCTGAGTTCGAGGCCAGCCTGGTCTACAA 4371
4372 AGTGAGTTCCAGGACAGCCAGGGCTACACAGAGAAATTCTATTCTTGGAAAAAAAACAACAAGGGAAAAAGAAAAAAAAAAAAAGAATC 4461
4462 TGAATGTGGAGGCTGGAGPGAYGGCTCAGTGGTTGTAACGCTACTGCTCTTCCTGAGGACCTGATTTCAATCCCCAGCAACACATGGTGG 4551
4552 CTCACAACCATATGCAATAGCATCCAATGCCTTCCTCGGGTGTGTTGGAAGACACCTACAGTGTATTCACCTACATGAAATAAATAAATC 4641
4642 TTRAGAAAAGTATGAGTGTGAACTCCTTGACCTGAGCAGTTCACTTGTGGGTAAATTTAAGGATTGGGAATATCTTAATATTGTGGAGGA 4731
4732 AATAAATGTAGGACCTTCCAGAATCTGTGTTTCTTGTGCTGAGTATGTCAGGTGGGGCAGGAGTCCTCTGGCATCTGCTTGTGAATGGGC 4821
4822 TGTGCCAGGTGGAGCTCAGTCCATGTGCACGTCCTTGCTGGTCTGGTCGATTCCTCAGCTATACCTCTGTGGTACTTGTCACTTAATAGA 4911
4912 GTCACTGGATGACAGACTCTGATGATGGTAACACGAAATGCCTTAGCCTTCTCTTATGACCGSACTATGAGCCCCGTAPACTGTACTAGG 5001
5002 CCTGTTGTGCAGCTGAAAAACTGACATTACCATTAAAAAAATTTCTGGAGAGTAGTGTGAACACCTTTAATCCCAACATTCAGGAGACAG 5091
5092 GCAGGAAAACCACTGATTTCTAGGCATCCTGGACACTCAGGGTGACACAGAGAAACCCTGTCTCCAAATTGGGTACCTGGGAGCAGGGTA 5181
5182 TCCCTGTGTCCAGCCTGACCATGCAGATTGTAGGCAGAAAGTAGATTTTGGGTTTTGGATTAGAAGAACGTTGCATGCAGGAAGCAAGGC 5271
5272 TTAAAGAGCCAAGCTAGTGGAGGCCTGGGAGGCAGAGCTGAGGGCAGGTTGACCTTTGTGGGCCCAGGTAATGAAGCTT 5350

Figure 3, pg. 2

Figure 4. Non-coding sequences of gene 17.3A. (A) Diagram showing the relationship between the third intron of the K^d gene and the 17.3A gene. Open boxes indicate exons, hatched boxes indicate Alu repetitive elements, and the solid regions in the introns (near 1803 in the K^d gene and 1265 in 17.3A) indicate simple sequences. (B) DNA sequence in the K^d and 17.3A genes at the borders of the 1093 bp sequence found in 17.3A. (C) DNA sequences bordering the 1130 bp sequence in the K^d gene.

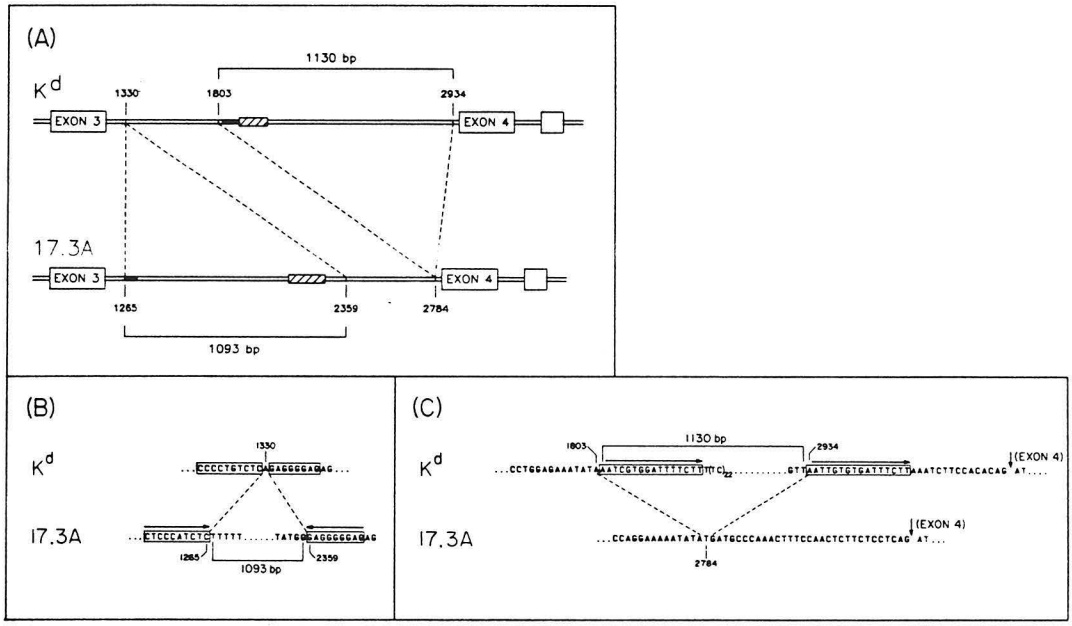
(A)
K^d
1130 bp
1330
1803
2934
EXON 3
EXON 4
17.3A
EXON 3
EXON 4
1265
2359
2784
1093 bp
(B)
K^d
1330
17.3A
1265
2359
1093 bp
(C)
K^d
1803
1130 bp
2934
(EXON 4)
17.3A
(EXON 4)
2784

Figure 4

Figure 5. Region of apparent gene conversion in 17.3A. (A) Exons 4 and 5 are shown as solid boxes, and introns are shown as open boxes. Numbers at the top indicate the percent nucleotide homology with the L^d gene in the bracketed areas. (B) Diagram showing the silent site comparisons made between 17.3A and the L^d gene.

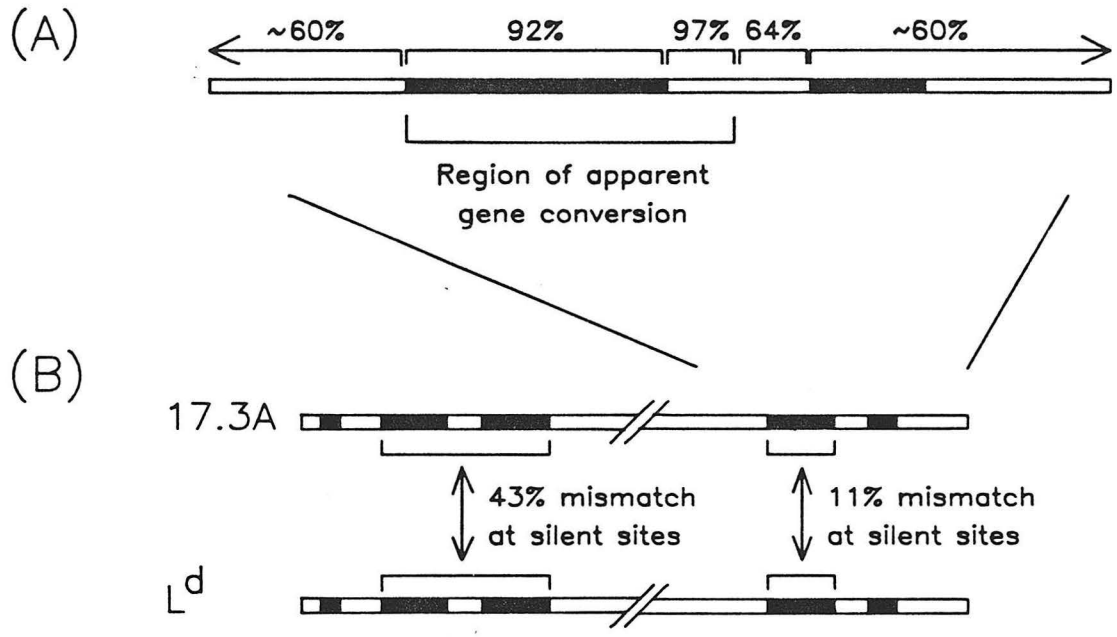
(A)
~60%
92%
97%
64%
~60%
Region of apparent
gene conversion
(B)
17.3A
43% mismatch
at silent sites
11% mismatch
at silent sites
L^d

Figure 5

Figure 6. Genomic southern blots with *Tla*-specific probes on DNA from four mouse strains. 10 μg (5×10^{-18} moles) of genomic DNA from the indicated strains was digested to completion with Hind III or Bam HI, electrophoresed through a 0.8% agarose gel in 1X TAE, blotted onto nitrocellulose and hybridized 12-18 h with 5×10^{5} cpm/ml of the indicated nick translated ($1\text{-}4 \times 10^{8}$ cpm/μg) probes. The final wash was in 1X SSC at 68°C. Hybridization markers are restriction enzyme digestion M13 subclones diluted to 10×10^{-18} moles per fragment. (A) Hind III digested DNAs hybridized with pTLA.4. (B) Hind III digested DNAs hybridized with pTLA.5. (C) Bam HI digested DNAs hybridized with pTLA.4. (D) Bam HI digests hybridized with pTLA.5. The bands in the Tla^{a} lane in (C) and (D) appear less intense than the other lanes because only 3-5 μg of DNA was loaded in those lanes. This inaccuracy was corrected in (A) and (B). All bands in the Tla^{c} lane do not hybridize with equal intensity. Those bands corresponding to gene 24.8 and the 24.8-like gene (genes T1 and T11 in Fig. 7; see Table II) are less intense in (B) and (D). The hybridizing region of 24.8 is only 85% homologous to pTLA.5, and would be expected to hybridize more weakly than the 100% homologous 17.3A gene (D. Fisher, unpublished results).

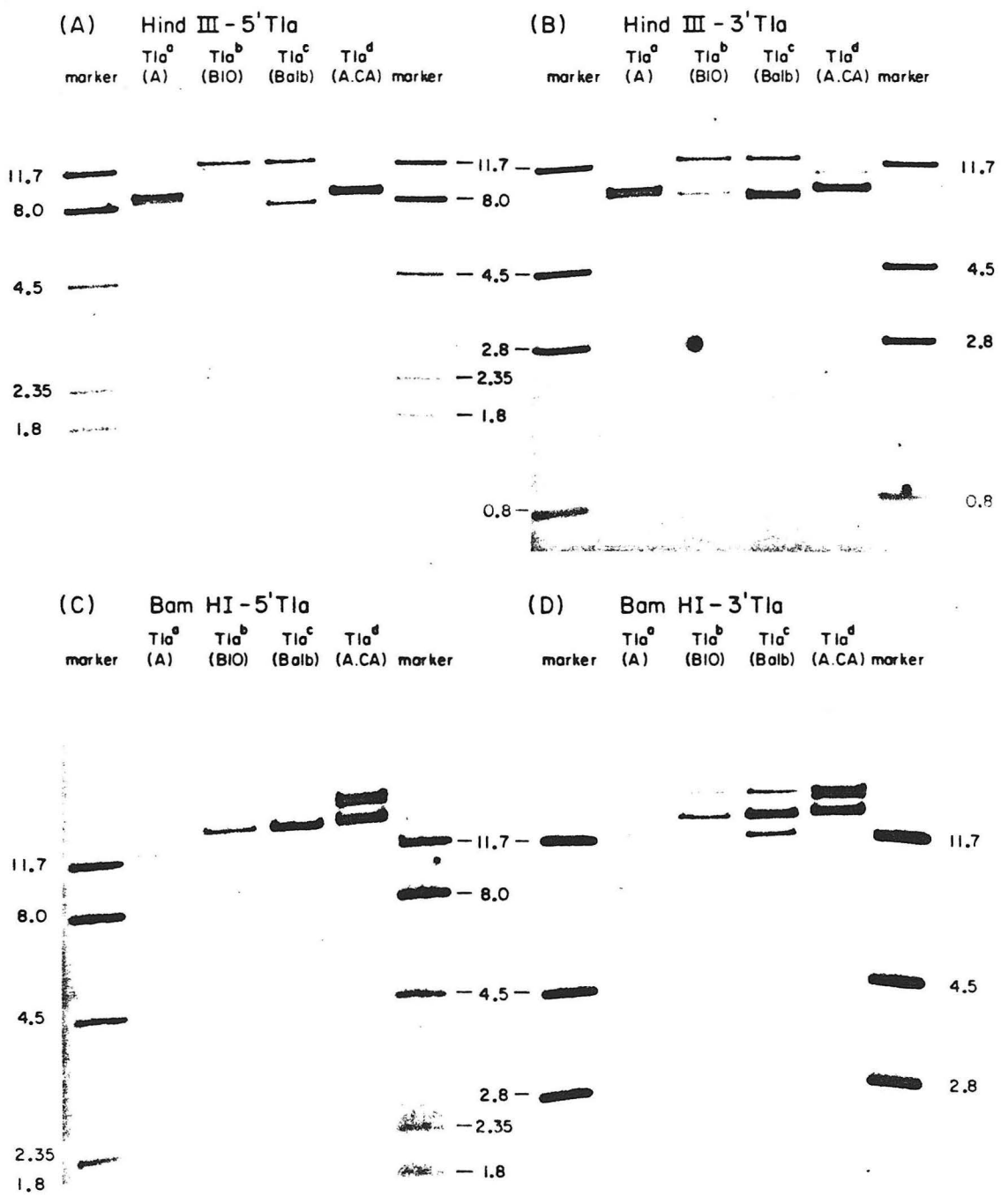
(A) Hind III - 5'Tla
marker
Tla^a (A)
Tla^b (BIO)
Tla^c (Balb)
Tla^d (A.CA)
marker
11.7
8.0
4.5
2.35
1.8
(B) Hind III - 3'Tla
marker
Tla^a (A)
Tla^b (BIO)
Tla^c (Balb)
Tla^d (A.CA)
marker
11.7
4.5
2.8
0.8
(C) Bam HI - 5'Tla
marker
Tla^a (A)
Tla^b (BIO)
Tla^c (Balb)
Tla^d (A.CA)
marker
11.7
8.0
4.5
2.35
1.8
(D) Bam HI - 3'Tla
marker
Tla^a (A)
Tla^b (BIO)
Tla^c (Balb)
Tla^d (A.CA)
marker
11.7
4.5
2.8

Figure 6

Figure 7. Class I *Tla* genes of the BALB/c mouse. (A) Restriction map and organization of gene cluster A. Regions hybridizing with class I cDNA probes (11) are indicated as thin open boxes. 5' (pH-2III) or 3' (pH-2IIa) hybridizing regions are indicated above the boxes, and the transcriptional orientation (determined by hybridization with the above probes) is indicated with an arrow if it is known. The location of the sequences hybridizing with pTLA.5 are indicated by an *, while those hybridizing with pTLA.4 are indicated by an X. The genes are numbered above the 5' and 3' designations. Open, solid and hatched bars at the top indicate regions that by their restriction maps appear to be related, and probably represent large duplicated regions. The regions covered by a representative group of cosmid clones, including the three artifactual constructs, are shown below the restriction map. Numbers in parentheses refer to the cluster designation in the Steinmetz et al. (13) paper. (B) Restriction map and organization of gene cluster B. (C) Restriction map for cluster 10.

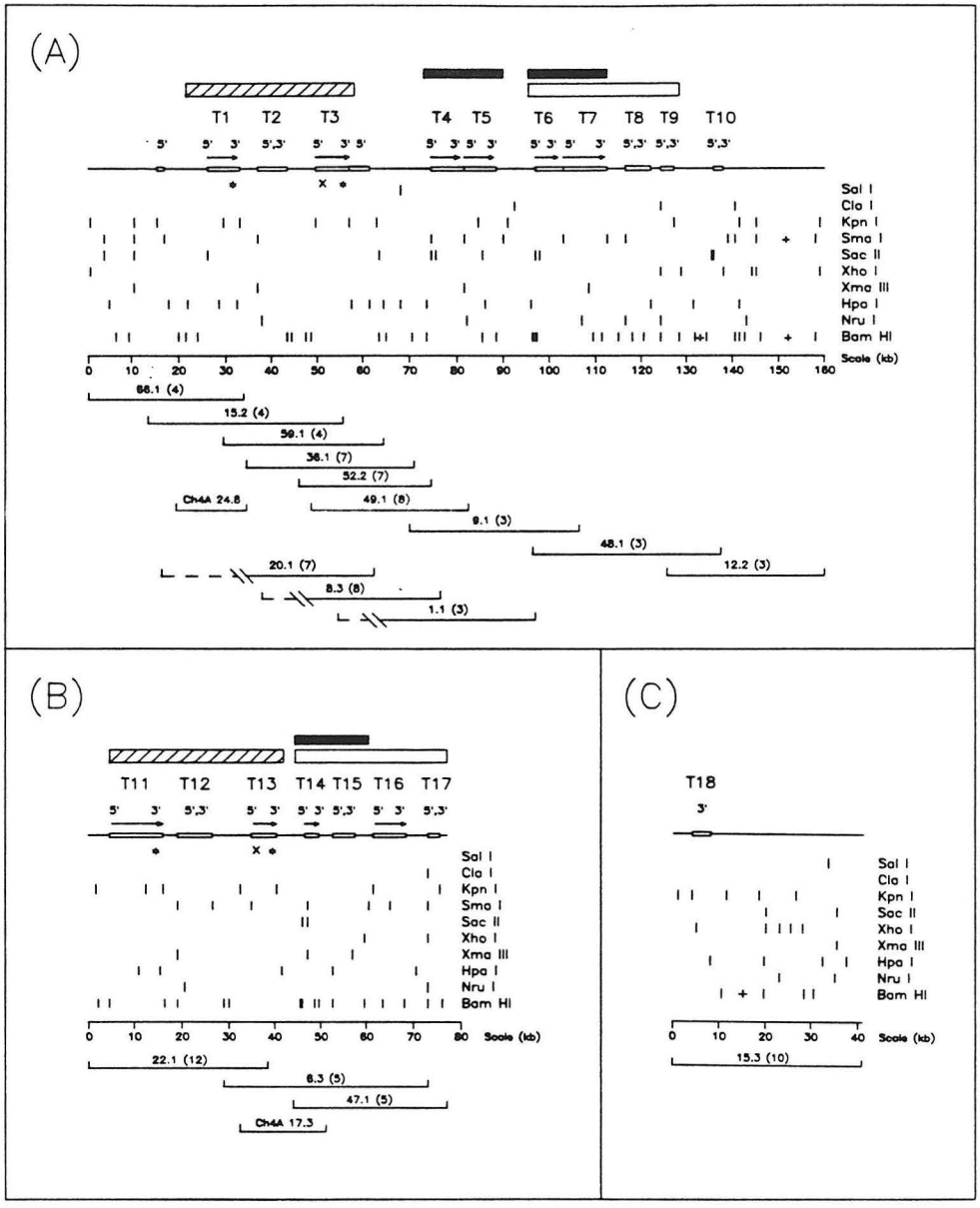
(A)
T1 T2 T3 T4 T5 T6 T7 T8 T9 T10
Sal I
Cla I
Kpn I
Sma I
Sac II
Xho I
Xma III
Hpa I
Nru I
Bam HI
Scale (kb)
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160
68.1 (4)
15.2 (4)
59.1 (4)
38.1 (7)
52.2 (7)
Ch4A 24.8
49.1 (8)
9.1 (3)
48.1 (3)
12.2 (3)
20.1 (7)
8.3 (8)
1.1 (3)
(B)
T11 T12 T13 T14 T15 T16 T17
Sal I
Cla I
Kpn I
Sma I
Sac II
Xho I
Xma III
Hpa I
Nru I
Bam HI
Scale (kb)
0 10 20 30 40 50 60 70 80
22.1 (12)
6.3 (5)
47.1 (5)
Ch4A 17.3
(C)
T18
Sal I
Cla I
Kpn I
Sac II
Xho I
Xma III
Hpa I
Nru I
Bam HI
Scale (kb)
0 10 20 30 40
15.3 (10)

Figure 7

Figure 8. Comparison of Tla^c (BALB/c) and Tla^b (C57BL) Tla region class I genes. Tla^c genes are shown as open boxes and Tla^b genes as solid boxes. Cleavage sites for the enzymes Kpn I (K) and Cla I (C) are indicated.

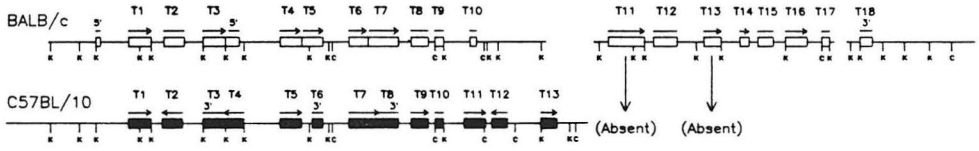

Figure 8

Chapter 4

DNA SEQUENCE OF A *Tla* REGION PSEUDOGENE*

By DOUGLAS A. FISHER, MARIT PECHT†, and LEROY HOOD

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

†Present address: Department of Cell Biology, The Weizmann Institute of Science, Box 26, Rehovot 76 100, Israel

Running title: SEQUENCE OF A *Tla* REGION PSEUDOGENE

*Supported by grants from the National Institutes of Health

Introduction

Thymus leukemia (TL) antigen, encoded at the *Tla* locus of mouse chromosome 17, is a cell surface glycoprotein expressed on the thymocytes and activated lymphocytes of TL^+ mouse strains, and on some leukemias of both TL^+ and TL^- mouse strains (1-3). The nature of the controls restricting TL antigen expression are unknown, as is the significance of TL antigen expression on leukemic cells and its role in the leukemic phenotype. Structurally, TL antigen is a ~45,000 dalton polypeptide that is closely related to the transplantation antigens, Qa-2,3 antigen, and Qa-1 antigen encoded at the closely linked *H-2*, *Qa-2,3*, and *Qa-1* loci, respectively (2, 4-6). All these antigens (denoted class I molecules) have similar molecular weights, associate non-covalently with β_2-microglobulin, and are encoded by genes (class I genes) that cross hybridize with nucleic acid probes (2, 7-9). Only the H-2 antigens, which are important regulatory molecules for cytotoxic T lymphocytes (10), have a known function, but the presence of TL and Qa antigens on lymphoid cells suggests that they too may be regulatory molecules in the immune system.

Class I gene sequences have been cloned (11), and the genomic DNA sequence of numerous genes determined (12-18). The isolation and analysis of cosmid clones containing class I genes showed that the approximately 30 genes (many more than the number of known gene products) are located in large gene clusters (7, 19) and that most genes come from the *Qa-2,3* or *Tla* regions (8). A total of 18 genes are encoded in the BALB/c *Tla* region (20). The *Tla* genes have been named T1-T10 in the Tla^b C57BL mouse by Weiss *et al.* (19) and T1-T18 in the Tla^c BALB/c mouse by Fisher *et al.* (20). We propose that a superscript indicating the *Tla* allele be added to the gene names to avoid confusion (Fig. 1). To identify genes encoding serologically defined products, the cloned genes were transformed into mouse L cells by DNA-mediated gene transfer and the resulting

transformants assayed with monoclonal antibodies (MAbs) against known class I determinants (9). By this method, two BALB/c (*Tla*^{c}) genes encoding TL determinants were identified, on clones λ17.3 and λ24.8.

The gene encoding a serologically defined TL antigen from clone λ17.3 (denoted gene 17.3A or T13^c) has been completely sequenced (20). The T13^c gene is similar in intron-exon organization to other class I genes with the exception of having its cytoplasmic domain encoded by one exon (exon 6) rather than in three exons (exons 6, 7, and 8). The nucleotide sequence of T13^c has diverged considerably (~25%) from genes in the *H-2* complex. Recently, using a probe from the T13^c gene, two groups have isolated and sequenced a functional *Tla* gene from the C57BL mouse (21, 22). This gene, the T3^b gene, is extraordinarily similar to the T13^c gene (95.5% nucleotide homology), but has a cytoplasmic domain that is 23 amino acids longer than that in T13^c.

In this report, we present the nucleotide sequence of the class I gene on clone λ24.8, which we denote gene T1^c. In spite of its earlier assignment, the T1^c gene cannot encode a TL antigen, or any other functional class I molecule, because it contains numerous stop codons and frameshift mutations in the coding regions.

Materials and Methods

Materials. Restriction endonucleases and T4 DNA ligase were obtained from New England Biolabs, Beverly, MA. *E. coli* strain JM103, phage M13mp8, and the large fragment of *E. coli* DNA polymerase I were from Bethesda Research Laboratories, Gaithersburg, MD. Deoxynucleoside triphosphates and DNAase I were obtained from Sigma Chemical Co., St. Louis, MO., dideoxynucleoside triphosphates and the 15 nucleotide sequencing primer from Collaborative Research, Waltham, MA., and α-[^{32}P] deoxynucleoside triphosphates were from Amersham Corp., Arlington Heights, IL.

Methods. Commonly used recombinant DNA procedures such as gel electrophoresis, Southern blotting, nick translation, growth of plasmids and phage, and restriction nuclease digestions followed previously described protocols (24, 25).

DNA Sequence Analysis. The sequencing reactions for the chain termination or chemical degradation methods were carried out according to previously published procedures (25, 26). Overlapping subclones for chain termination sequencing in phage M13mp8 were generated using a DNAase deletion subcloning method similar to several published methods (27-29).

Results and Discussion

Gene $T1^c$ is a Pseudogene by Numerous Criteria. Clone λ24.8 was originally isolated by virtue of its hybridization to class I cDNA probes (11; K. Moore, Caltech, unpublished results). Since λ24.8 transformed mouse L cells to the TL^+ phenotype (9), and therefore presumably contained a gene encoding TL antigen, it was restriction mapped (Fig. 2A), subcloned (Fig. 2B), and sequenced using the strategy in Fig. 2C. The DNA sequence is shown in Fig. 3.

In general, class I genes are composed of eight exons that correlate with protein domains: exon 1 encodes the leader peptide, exons 2, 3, and 4 code for the α1, α2, and α3 external domains, exon 5 codes for the hydrophobic transmembrane peptide, and exons 6, 7, and 8 encode the cytoplasmic domain (12-18). In the recently sequenced functional *Tla* gene ($T13^c$ in Fig. 1), this arrangement is the same except the cytoplasmic domain is encoded by a single exon, exon 6 (20).

The $T1^c$ gene contains sequences related to all eight class I exons, but an examination of these sequences reveals numerous in-frame termination codons and frameshift mutations (Table I). The only ATG initiation codon located near the leader peptide is followed two codons later by a termination codon. This initiation

codon is out of frame with the exon 1 reading frame. Near the 3' end of exon 1 (between bases 869 and 870, labeled "I" in Fig. 3) a single base deletion disrupts the reading frame again and causes the splice to occur out of frame. This leads to premature termination in exon 2 at amino acid 4 (bases 1085-1087, underlined in Fig. 3). Even assuming a normal splice to exon 2, a termination codon is encountered at amino acid 27 (bases 1153-1155). Since there are ten mutations that would render gene $T1^C$ nonfunctional, and since five of these have been confirmed by second strand sequencing by the method of Maxam and Gilbert (26), we conclude that the gene is a nonfunctional pseudogene that could not by itself encode TL antigen or any other class I molecule.

Large Blocks of $T1^C$ Sequence Are Closely Related to the Functional $T13^C$ Gene, While Other Parts Are Unrelated. The $T1^C$ and $T13^C$ genes are clearly related because a transmembrane/cytoplasmic region probe, pTLA.5, subcloned from $T13^C$, hybridizes to gene $T1^C$ (20). Probe pTLA.5 is specific for TL^+ mRNA and hybridizes with only four of the thirty-three BALB/c class I genes ($T1^C$, $T3^C$, $T11^C$, and $T13^C$) (20). However, a similar TL^+-RNA-specific probe subcloned from exons 2 and 3 of gene $T13^C$, pTLA.4, does not hybridize to gene $T1^C$ (20), indicating that this part of the gene is less closely related to $T13^C$ than the 3' region.

Comparison of the DNA sequence of genes $T1^C$ and $T13^C$ (Table III) confirms the pattern of hybridization for pTLA.5 and pTLA.4 just noted. The transmembrane and cytoplasmic exons have 85% nucleotide homology, while exons 1-3 have only 64% homology. Indeed, the first three exons of gene $T1^C$ are slightly more closely related to a gene from the *H-2* region, the *H-2D*^{d} gene (Table II).

Alignment of the $T1^C$ and $T13^C$ genes have shown that the 3' part of the third intron and all sequences 3' to the third intron have 85% homology at the nucleotide level, while the 5' part of the third intron and all introns and flanking regions upstream are unrelated (<40% nucleotide homology) (Fig. 4A). There is a

sharp boundary between the related and unrelated sequences, and this occurs precisely at a B2 Alu-like repetitive element (30) present in the third intron of the $T13^C$ gene (Fig. 4A, B).

Gene $T1^C$ appears to be a hybrid gene composed of sequences closely related to a functional Tla^C gene in its 3' half, and sequences from another class I gene in its 5' half. This gene was probably generated by a recombinational event like unequal crossover, with the recombination point at an Alu-like repetitive element. Since Alu-repeat elements have some of the characteristics of transposons, the B2 repeat might have catalyzed a recombinational event by a transposon-like insertion. On the other hand, there might have been Alu-repeats in both $T1^C$ and $T13^C$, one of which was later deleted, and a recombination occurred by simple crossover between homologous sequences.

The Transformation Result with Clone λ24.8 is not Easily Explained by the $T1^C$ Gene Sequence. The sequence of gene $T1^C$ shows that the TL^+ transformation result (9) cannot be simply explained by the expression of the introduced gene. Three major possibilities should be considered: 1. The transformation result was an artifact. 2. A small part of the $T1^C$ gene recombined with a host cell class I gene and transferred the information for TL antigen epitopes recognized by the anti-TL mAbs to an expressed gene. 3. Sequences within or flanking gene $T1^C$ brought about the induction of a serologically reactive *Tla* gene in the host L cell.

The second possibility is unlikely because the epitopes recognized by the anti-TL mAbs are located in the α1 and α2 protein domains encoded by exons 2 and 3—precisely the part of the $T1^C$ gene that is not related to the functional $T13^C$ gene. When a hybrid gene composed of exons 1-3 of gene $T13^C$ and exons 4-8 from L^d gene is transformed into L cells, the resulting transformants are TL^+ and L^{d-} (S. Hunt, Caltech, unpublished results). In the reciprocal experiment (L^d exons 1-3 + $T13^C$ exons 4-6), the transformants are L^{d+} and TL^-. The second and

third exons of $T1^C$ not only have low homology to the exons encoding TL determinants, but these $T1^C$ exons are also riddled with stop codons and frameshift mutations that would make it less likely to transfer any functional gene segments.

The third possibility, that of host gene activation, cannot be ruled out, because the anti-TL mAbs bind to TL molecules from both the donor (Tla^C) strain and the host L cell (Tla^b) (9). However, the first possibility, that of an artifact, looms as the most likely explanation for the transformation result of λ24.8. The finding that the $T1^C$ and $T13^C$ genes are two members of a four-member subfamily of *Tla* genes is interesting, but at least three groups have been unable to repeat the λ24.8 transformation result under conditions that easily detected λ17.3 ($T13^C$) transformants (S. Hunt, Caltech, Pasadena, CA; F.-W. Shen, Mem. Sloan Kettering Cancer Center, NY, NY; S. Nathanson, Albert Einstein College of Medicine, Bronx, NY, unpublished results). Gene $T1^C$ does not encode TL antigen in any simple way, and more complicated explanations (recombination or gene activation) seem less likely than that of the initial $T1^C$ transformation result being an artifact.

Summary

We have determined the DNA sequence of a BALB/c *Tla* region class 1 gene that had been previously identified as encoding TL antigen by DNA-mediated gene transfer. Analysis of the DNA sequence shows, however, that this gene, the $T1^C$ gene of Tla^C, could not encode the TL antigen or any other functional class 1 molecule due to the presence of numerous stop codons and frameshift mutations in the coding regions. This result suggests either that the earlier transformation data were in error, or that the clone containing the $T1^C$ gene may also contain sequences that induced the resident *Tla* gene in the transformation host L cell.

The $T1^C$ gene is structurally related to the previously sequenced $T13^C$ gene that encodes a serologically defined TL antigen. The 3' half of the $T1^C$ gene including exons 4, 5, 6, and the 3' untranslated region has about 85% nucleotide homology (including introns) with the corresponding parts of the $T13^C$ gene; however, the 5' half of the $T1^C$ gene has little homology with the $T13^C$ gene. There is a sharp line of demarcation between the homologous and non-homologous regions, and this border occurs precisely at a B2 Alu-like repeat sequence present in the T13 gene.

Acknowledgement

We thank Karyl Minard and Debbie Maloney for expert technical assistance.

References

1. 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. Nat. Cancer Inst.* **31**:977.

2. Flaherty, L. 1980. The Tla region antigens. *In* Role of the Major Histocompatibility Complex in Immunology. M. E. Dorf, editor. Garland Press, New York, pp. 33-57.

3. Cook, R. G., and N. F. Landolfi. 1983. Expression of the thymus leukemia antigen by activated peripheral T lymphocytes. *J. Exp. Med.* **158**:1012.

4. Vitetta, E. S., and J. D. Capra. 1978. The protein products of the murine 17th chromosome: Genetics and Structure. *Adv. Immunol.* **26**:147.

5. Yokoyama, K., E. Stockert, L. J. Old, and S. G. Nathanson. 1981. Structural comparisons of TL antigens derived from normal and leukemia cells of TL^+ and TL^- strains and relationship to genetically linked H-2 major histocompatibility complex products. *Proc. Nat. Acad. Sci. USA.* **78**:7078.

6. McIntyre, K. R., U. Hämmerling, J. W. Uhr, and E. S. Vitetta. 1982. Structural analysis of thymus leukemia (TL) antigens in the mouse. *J. Immunol.* **128**:1712.

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

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

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

10. 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.

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

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

13. 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.

14. Evans, G. A., D. H. Margulies, R. D. Camerini-Otero, K. Oato, and J. G. Seidman. 1982. Structure and expression of a major histocompatibility antigen gene H-2L^d. *Proc. Nat. Acad. Sci. USA* **78**:1994.

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

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

17. Mellor, A. L., E. H. Weiss, M. Kress, G. Jay, and R. A. Flavell. 1984. A nonpolymorphic class I gene in the major histocompatibility complex. *Cell.* **36**:13.

18. Sher, B. T., R. Nairn, J. Colligan, and L. Hood. 1985. DNA sequence of the mouse D^d gene. *Proc. Nat. Acad. Sci. USA* **82**:1175.

19. 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.

20. 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.

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

22. Pontarotti, P. A., H. Mashimo, R. A. Zeff, D. A. Fisher, L. Hood, A. Mellor, R. Flavell, and S. G. Nathenson. 1985. Conservation and diversity in the class I genes of the major histocompatibility complex: Sequence analysis of a Tla^b gene and comparison with a Tla^c gene. *Proc. Nat. Acad. Sci. USA*, submitted.

23. Davis, R. W., D. Botstein, and J. R. Roth. 1980. Advanced Bacterial Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

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

25. 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.

26. Maxam, A. M. and W. Gilbert. 1980. Sequencing end labeled DNA with base specific chemical cleavages. *Meth. in Enzymol.* **65**:499.

27. Frischauf, A. M., H. Garoff, and H. Lehrach. 1980. A subcloning strategy for DNA sequence analysis. *Nucleic Acids Res.* **8**:5541.

28. Hong, G. F. 1982. A systematic DNA sequencing strategy. *J. Mol. Biol.* **158**:539.

29. Barnes, W. B., M. Bevan, and P. H. Son. 1983. Kilo-sequencing: Creation of an ordered nest of asymmetric deletions across a target sequence carried on phage M13. *Methods in Enzymol.* **101**:98.

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

31. Davis, M. M., K. Calame, P. W. Early, D. L. Livant, R. Joho, I. L. Weissman, and L. Hood. 1980. An immunoglobulin heavy chain gene is formed by at least two recombinational events. *Nature (Lond.).* **283**:733.

Table I

Summary of Mutations Rendering $T1^C$ *Nonfunctional*

Region		
Leader	1.	The only nearby ATG initiation codon, which is not even in frame with the leader, is followed two codons later by a stop codon.
	2.	Frameshift (1 bp deletion at amino acid [AA] -3*) causes an out of frame splice. Stop codon encountered at AA 4.
α1	3.	In frame stop codon at AA 27.
	4.	Frameshift (20 bp deletion at AA 40-46). Stop codon at AA 69.
α2	5.	Stop codon at AA 97.
	6.	No GT splice donor sequence. Stop codon 13 codons into intron.
α3	7.	Frameshift (5 bp deletion at AA 194-195). Stop codon at AA 196.
	8.	Frameshift (1 bp insertion at AA 256). Stop codon at AA 264.
	9.	Frameshift (14 bp deletion at AA 267-271). Stop codon 1 codon into intron.
TM	10.	Longest stretch of uncharged amino acids is 14—too short to span the membrane.

*Amino acid numbering refers to the aligned $T13^C$ sequence (see ref 20). The location of the various stop codons, deletions, and the one insertion are shown in the $T1^C$ gene sequence in Fig. 3.

Table 2

Exon by Exon Nucleotide Homologies

of $T1^c$ *with BALB/c Tla and H-2 Genes*

	$T13^c$	D^d
Exon 1 (leader)	52%	62%
Exon 2 (α1)	64%	66%
Exon 3 (α2)	67%	69%
Exon 4 (α3)	88%	86%
Exon 5 (TM)	85%	46%
Exon 6 (Cyto)	85%	63%

Figure 1. Class I genes of the *Tla* region in BALB/c (Tla^c) and C57BL/10 (Tla^b) mice. Gene names after Weiss *et al.* (19) and Fisher *et al.* (20) with the exception of the added superscripts to indicate the *Tla* allele. Cleavage sites for the restriction enzymes Cla I (C) and Kpn I (K) are shown.

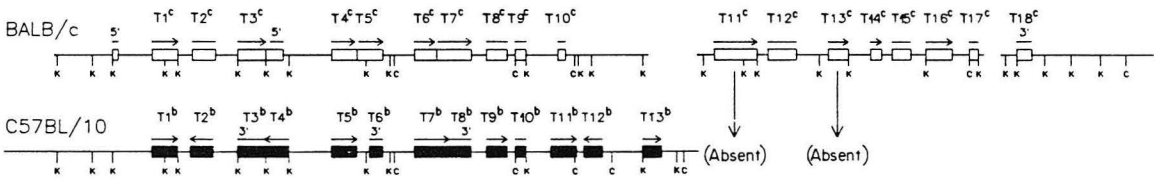

Figure 1

Figure 2. Restriction map and sequencing strategy for class I gene $T1^C$, from clone λ24.8. A) Restriction map of the eukaryotic insert of λ24.8. EcoRI sites indicated by an asterisk (*) result from the addition of EcoRI linkers during the construction of the library (31), and are not present in the genome. B) Restriction map of subclone pTLA.8, a pBR322 subclone derived from λ24.8. C) Sequencing strategy for gene $T1^C$. Regions homologous to the first six exons of class I genes are shown as solid boxes. Introns and flanking sequences are shown as open boxes, the 3' untranslated region by a hatched box, a B1 Alu repeat as a dotted box, and a 500 bp gap in the sequence by a dashed line. The cleavage sites for selected restriction enzymes are shown, as is the first polyA addition signal (AATAAA) in the 3' untranslated region. Individual sequencing runs done by the chain termination method (25) are shown as straight arrows and those done by the chemical degradation method (26) as wavy arrows. The location of sequences hybridizing to the 3' *Tla*-specific probe pTLA.5 from the $T13^C$ gene are shown.

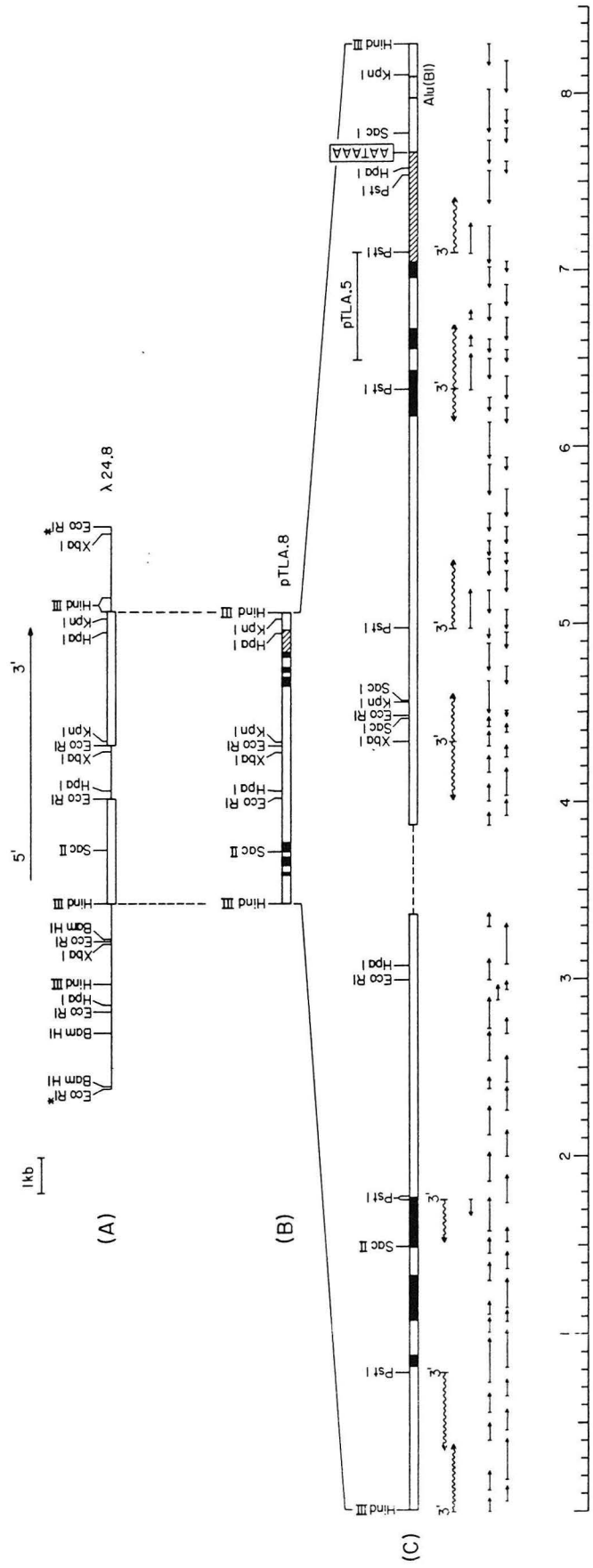
λ 24.8
pTLA.8
pTLA.5
AATAAA
Alu(BI)
1kb
(A)
(B)
(C)
5'
3'
Hind III
Sac II
Pst I
Xba I
Kpn I
Hpa I
Eco RI
Bam HI
Sac I

Figure 2

Figure 3. DNA sequence of gene $T1^C$. Regions homologous to the first six class I exons are underlined, while the ATG initiation codon, (bases 795-797), the termination codon (bases 7040-7042), and the polyA addition site (bases 7663-7668) are boxed. Solid triangles show the locations of the following deletions relative to the $T13^C$ sequence: 1 bp (I), 20 bp (II), 5 bp (III), and 14 bp (IV). An asterisk (*) shows the location of a one base pair insertion in the fourth exon. The stop codons noted in Table I are double underlined. Ambiguous base code: N = A, G, C, T; P = A, G; Y = C, T; S = G, C; R = A, T.

Figure 3, pg. 1


4501 GAGGGATTGAGTGTGGGCTGTTGTAAATATTTCTACCCCAGACAGGGATGGTACCCAGAGCTCAGATGTGTCTCCCATAGATGTAAAGGC 4590
4591 AACACCCTAATGATCAAATCATCCGTTGTGACAGATTTGGACATGCCTGGTTCACAGAAACGTGCAGAATTTAGACCTTTCAGTGAGTGT 4680
4681 GTGGGTGGTTGGGATATGTCTTCACGGAAATGTCTGTGAATCTTCTCCTACAGCATGAAGCCTGGCATGAACTGAGTGACAGAAAATGTG 4770
4771 TTGGGGTCTCTCCTGTGACATCCAGAGCACTCAGTTCTCTTTCAGCAACAGTGTCTGAATTTCCCTGTGATCCTATGAGCACAATGTGAA 4860
4861 GAACTGTGAAGCCCAATCTACGCCTGCACTCTAGGACCCTCTCCCTGCACTGCCTGTGTTCCCTTCCACGGAGGAGACATCTGCATCCTG 4950
4951 TCAGCTCCATGCTGCCCTGACTGCAGCTCCTCACTTCCACACTGAGAATAAGAATATGCATGTGAACTTGATTGTTCACACCCTTGATCT 5040
5041 GACAGGATGATTGACAGGTGTATTAAAGGACTGAGAATACTTAGAGTTTTGGGGAAGAAATAAATGGCAGCATGGAGAACCTTCTGGAAT 5130
5131 CTGTGTTCTTTATGCTGAGTGTGTCTGGTGGGAAAGAGGAGTCTGTGGGAGCTGAGTGTGAATAGGGCACAACCCAGGTAGGGCTCAGTT 5220
5221 GGTTGTAGGCTTTGATGTGGTCATTCCTCAACTCGGGTCACCTTTTGGCTGTGTTCTCTTCATCAATCAATAGAATGACATGCTGACAGG 5310
5311 GTCTGTGANCCAGCATGGCCTTGTTGGTCTCCAGGATTATGAACAGCCAGGGCTGCTGTAACAGTGCAACCTGTGCTCTGACCTGGCTGT 5400
5401 GGTTCTCCTTCCTAGTTTCTGTGCAGGACCAGAGCTGTCCTAGTTCAAATGATGCGGGTCTTCTCCTTCCCTGGGCTCCTAGAGTAACAG 5490
5491 TCTGGACTCAGGGTCCAGAGATACTACCTGGGGCTGCCTNNCTGGTAGAGAAGGAGNNTCCTCCTGCTGCTCACAGAGCAACCZGGAANT 5580
5581 TAAAAAATAAAATATCAAATAAAACAGAGGAGATATCAAATAACACAATCTCTTCCCCTTCCTGATTTTAACAGAAACACTAGGTATCTC 5670
5671 TTCATTTCACTTACTGTTCCAGCCCTTTCCTTCTTGCTAGCATGAGGTAGGGGGAGATCACCCCCGTGTTTCCTGACTATATCAGAGACT 5760
5761 GGGGGCCCCGAGGGAGCACCCTTRTCTCAGGAGAGTTTGAAGTCTCGTCAGGATGAAAAGAGAGAAGGTTCCTGAAATAACCAAAGTTCC 5850
5851 CCCACTGTGTCTGCTTCCTTTAGAGCTATGGCTTCTCTCAGGCTTTGTTCTCTGCCCACACCCAACCTCTGTGACATTTGACTTCAGTGA 5940
5941 TGCTGAATCTCTGAGCACTNCCCAGACAGAAGCAGACCTCCCTTTTTTCCTTTGGGACACATAGCTGCTCCTGGGNANAAGTTCTGPNAA 6030
6031 TTTTCAGAGGATGTGTGTAAACCTACTGGGAAATTTGACCGTTGTTCCAGTNNTTTCCTTCCCAGGTAGCCTCACTGACACTTCTGGAGT 6120
6121 CTTCAGGAAGATTACAAGAGGTCCAAGCTATCCAAGTCTTCTCCTCAGATCCTCCAAAGACACACTCATCATCTCAGGCCTGAAGGTGAT 6210
6211 GTCACCCTGAGGAGCTGGGCCCTGGGCTTCTACCCTGCTGACATCACTGTGACCTGGCAGTTGAATGGAAAGGAACTGACCCATGACATG 6300
6301 GAGCTTGTGGAGACCAGGCCTGCAGGGGATGGGACCTTCCAGAAGTGGGCAGCTGTGGTGGTGCCTTCTGGGGAAGAGCAGAATATATAC 6390
6391 TTGGCATGTGTACCATGAAGGGCTGTCTGAGATGGGGTAAGGAGGGTGTGGGTGCAGAGCTGGGGTCAGGGAAAGCTGGAGCCTTCTGAA 6480
6481 GACCCTCAGCTGGTCAGAGTTGAGAGTTGCGATCATGTTCACTTTCCCCATGTCCTTACCCTTCTCTTCTCAGAGCCTCCTCAGTCCAGT 6570
6571 ATGCCCACCAGGACCATTGTTGGTGTTGTCCTTGGAGCTATGGTCATAGGTTTAAAGAGCAGAAATGTTATGATGTGGATGAGAAAGAAC 6660
6661 AAAAGTATGGAAGAGTCTGTGGCTAGGTGCCTTATGATTATAAACCAATACACATAGACTATATGCACAAGAATTTCCCATCTCTGGGTC 6750
6751 TATTTGCAGACACRCACCTTACAATAAAAGACAGATTCATAATGGTAATGGGAGAATCTGCAAAGGATTCTCAATGACCTCTCAGAGNNC 6840
6841 AGAGGGCAAGGCTCATGCTAAACACTACAGATCTCCAGGCTGGCCTTGGTTCATCCCTTGCATATAATGTCCATGGCCCAGAACTTCCTT 6930
6931 GGTCTCTCATATCTTTTTTTTTAAGAGGTGGAAACAGAGGCTATGACCCTGCTGCATGTCAGAATGTGAGGGAACAGTTATCCTTGAGGCA 7020
7021 CTCGGGGTGGAATCTTGGCTGAAGGTTCTTCCTCCAGCCCCACCATTGGTCTCCTGACAGTGTCTTGTTACTTCTACTGCAGGTAAGGAC 7110
7111 AGTTCCTGACTTATCTCTGGAGACTTGAAAAGGTGAGACCTGTGAGCCTGGGATTGGGGCTGAGGCAAACTGGAAAGGACTGAGTCACTG 7200
7201 GGCCTCTCAGATTTAGACCTTTGAATGAATGAGAGGAGGGCCTTTGAGATGTTTTCATATTGATTGCCCTGACATTGTTCTTCATCTTGG 7290
7291 GACAGCTGCCTTCTGGGACTGAATGACAAGTTCTGACTTCTGAACAGTGAAATCACTTTGGGCAGCGGTGTCCGAGGGCTTATTCTTTTC 7380
7381 CTGTGCACACAGTCGAGAGATCTGGAGAGCAGAGCCCACTCACACACACCAGAACTCTGTCCCTACACTGGCAGTCATCCCTTCCACAGC 7470
7471 TAATCTTCCTAGTGCAACAGACACTGGGGACATCTACATCCTATCAGTTCCATGCTGCCCTGAGCTGCAGCTCCTCACTTCCACACGTAA 7560
7561 CATAAGAATTTGAATGTTAACTTGATTGCTCAAATCTTTAACCTGAAGGGTTGAATAGTGAGTATATTAAGGGACTGAGACTACTTGGAN 7650
7651 TGCTGTGGAGGAAATAAATGGCAGCATGGAGGACCTTCCAGAATCTGTGTTTCTTGTGCTGAGTGTGTCAGGTGGGGCAGGAGTCCTCTG 7740
7741 GCATCTGTTTGTGAATGGGGCTGTGCTAGGTGGAGCTCAGTCCATGTGCACGTCCTTGCTGGTCTGGTCATTCCTCAGCTATACATCCTT 7830
7831 GGAACTTGTCACTTAATAGAGTCACTGGATGACAGACTCTGATCATAGGAACACAGCATGCCTTAGCCTTCTCTTATGACCAGGATTATG 7920
7921 AGCCCCTAGGACTGTACCAGGCCTGTTGTGCAGCTGCGAAACTGACATTACCAGTGAAGAAATTTCTGGAGAGTGGTCCTGAACACCTTT 8010
8011 AATCCCAACATTCAGGAGACAGACAGGAGAACCACTGATTTCCAGGCCATCCTCAACACTCAGGGTACACAGAAAAAACCCTGTCTCCAA 8100
8101 ATTGGGTACCTGGGATCAGGTATCCCTGTGTCCAGCCTGACCATGCAGATTGTAGGCAGAATGTAGATATTTGAGGTTTGGATTAGAAGA 8190
8191 ACAGTTGCATGCATGAAGCAAGGCTTAAAGACCAACTATGAGCCTGGAGCAGAGCTGAGGGCAGGTTGACCTTTGTGGGCCCCGGTAATG 8280
8281 AAGCTT 8286
EXON 4
III
IV
EXON 5
EXON 6

Figure 3, pg. 2

Figure 4. Comparison of genes $T1^C$ and $T13^C$. A) Alignment of the $T1^C$ and $T13^C$ genes showing the similarity of the 3' part of the genes (note the restriction maps), and the B2 Alu repeat where homology breaks off in the third intron. Parts of the genes (introns, exons, and Alu repeats) are shown as in Fig. 2. B) Detail of the DNA sequence at the breakpoint of homology between the two genes showing its proximity to the B2 Alu repeat. An arrow is above an eight base pair direct repeat that presumably forms the border of the Alu repeat.

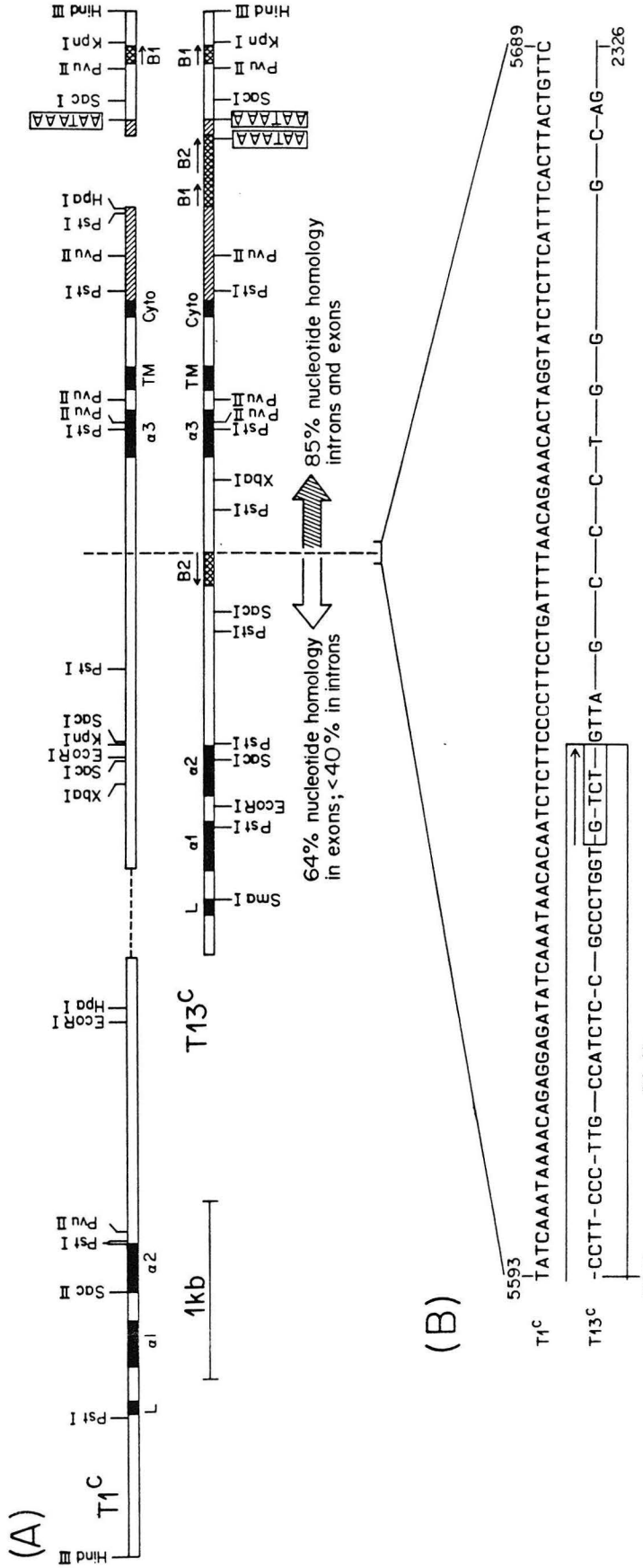
(A)
T1C
T13C
1kb
64% nucleotide homology in exons; <40% in introns
85% nucleotide homology introns and exons
(B)
5593
5689
2326
2230
B2 Alu

Figure 4

Conclusions

Structure and Evolution of H-2 Genes. This study shows the DNA sequence of cDNA clones encoding mouse transplantation antigens. The sequence convincingly establishes the homology of immunoglobulin (Ig) and transplantation antigen genes. Previous protein sequencing studies had suggested that Ig and H-2 molecules were related, and this raised the possibility that H-2 molecules, like Ig molecules, might have extensive somatic variability generated from DNA rearrangements. In addition, to explain the appearance of inappropriate H-2 serological determinants on some tumor cell lines, it was suggested that a mouse might have genes for all the hundreds of *H-2* alleles and that ordinarily only one gene is expressed.

Both of these hypotheses were disproved in Southern blotting experiments using the H-2 cDNA clones as probes. There are too few genes, as estimated from the number and intensity of bands on a genomic Southern blot, to encode all the H-2 alleles. The failure to detect any restriction enzyme fragment polymorphism between germline and liver DNA with the H-2 probes showed that H-2 genes do not employ the DNA rearrangements associated with the generation of antibody diversity. Instead, the similarity of H-2 molecules to Ig molecules is limited to the constant regions encoding such effector functions as complement fixation, raising the interesting possibility that transplantation antigens may at some time carry out similar functions.

Structure and Evolution of Tla Genes. The complete DNA sequence of a gene encoding a serologically defined TL antigen establishes the structural relatedness of H-2 and TL antigens. Previously, this relationship was no more detailed than their similar molecular weights, association with β_2-microglobulin, and a ~20% peptide map homology. The *Tla* gene is less closely related to *H-2* genes than *H-2* genes are to each other or to *Qa-2,3* region genes. Although this

means that the *H-2* and *Tla* genes diverged long ago, there is an apparent recent gene conversion including the fourth exon that could have been transferred to the *Tla* gene from an *H-2* gene. However, this gene conversion is probably of no functional significance—a product of the already high nucleotide homology of the fourth exon encoding the β_2-microglobulin binding domain.

Examination of the 18 cross hybridizing class I genes of the BALB/c *Tla* region indicates that these genes have evolved by large-scale duplications. More importantly, comparison of the Tla^c genes with those from Tla^b provides a possible explanation for the TL^- phenotype of Tla^b mice.

Tla Phenotypes and Structure: Antigenic Modulation. Part of the structure of a TL antigen deduced from the DNA sequence may be a clue to the antigenic modulation of TL molecules by antibody. In this process, TL antigens (unlike H-2 antigens) are removed from the cell surface by a patching and capping mechanism—events presumably involving contact between the TL antigen and the cytoskeleton. The cytoplasmic domain of the TL antigen is radically different from H-2 molecules, and this may reflect different interactions with cytoskeletal components.

This hypothesis can be tested by constructing hybrid *Tla/H-2* genes and transforming them into mouse L cells to locate the structural component responsible for the antigenic modulation effect. For example, one could construct an L^d antigen having the cytoplasmic domain of a TL antigen, and also construct a TL antigen with the cytoplasmic domain of the L^d molecule. Is the former now modulated by anti-L^d antibody, and is the latter, although it has all its external domains from the TL antigen, no longer modulated by anti-TL antibody? Studies such as these should locate the part of the TL molecule responsible for antigenic modulation.

The Phenotypes and Structure: Expression in TL⁻ (Tla^b) and TL⁺ (Tla^c) Mice. Examination of the organization of *Tla*^{b} and *Tla*^{c} class I genes suggests that the TL⁻ phenotype of *Tla*^{b} thymocytes is probably due to the deletion of a *Tla* structural gene rather than the alteration of a regulatory gene. The T3^b gene of *Tla*^{b} mice encodes the TL antigen expressed on their leukemia cells (1), while *Tla*^{c} mice have the allele of the T3^b gene, T3^c, as well as a gene encoding thymocyte TL antigen, T13^c. There is no allele of the T13^c gene present in *Tla*^{b} mice.

Are the TL antigens expressed on *Tla*^{c} leukemia cells the product of the T3^c gene, the T13^c gene, or both? This question cannot be definitively answered yet, but evidence points to both being expressed. The only TL antigenic determinant detectable on *Tla*^{c} thymocytes is TL.2, and therefore these cells are denoted TL.1$^-$2$^+$3$^-$4$^-$. L cells transformed with the T13^c gene are TL.1$^+$2$^+$3$^-$4$^-$ (2). *Tla* leukemia cells express new determinants, as they are either TL.1$^+$2$^+$3$^-$4$^-$ or TL.1$^+$2$^+$3$^-$4$^+$ (3). One report showed that the failure to detect TL.1 and TL.4 on *Tla*^{c} thymocytes may be due to the 20-fold lower level of expression of TL antigen in *Tla*^{c} thymocytes (4). It may also be that some weakly cross-reacting determinants (TL.4?) may only be detectable on *Tla*^{c} thymocytes by the immunochemical technique used and not by serological methods that assay specificity by cell lysis with antibody and complement.

A recent study showed that *Tla*^{c} leukemia cells express two TL polypeptide chains with similar, but not identical, peptide maps (5). One of these, an ~47 kDa glycoprotein, is identical with the TL antigen expressed on *Tla*^{c} thymocytes, and the other is an ~50 kDa glycoprotein. These differences are not due to glycosylation. It is not certain whether the two polypeptides are products of different *Tla* genes or the product of differential splicing from a single gene; however, the former seems more likely for three reasons. First, the only TL antigen other than that on *Tla*^{c} leukemias to have the TL.1$^+$2$^+$4$^+$ phenotype is the T3^b molecule. The

$T3^b$ and $T3^c$ genes are very closely related, even more so than the $T3^b$ and $T13^c$ genes, which are known to have 97% nucleotide homology (1). Therefore, it is reasonable to suspect that the $T3^c$ molecule would also have the TL.4 determinant. If this were the case, then Tla^c leukemias expressing the $T3^c$ gene would also express a new antigenic determinant, TL.4, as observed. Second, the $T3^b$ molecule (and probably the $T3^c$ molecule) is 2.3 kDa larger than the $T13^c$ molecule due to an additional 23 amino acids at its carboxy terminus. This is in agreement with the larger size of the second polypeptide. Third, there are no convincing alternative splices of the $T13^c$ gene that would produce a larger polypeptide.

The following is the pattern of TL expression deduced from the above data. The $T13^c$ gene is expressed on Tla^c thymocytes and encodes a protein that is $TL.1^+2^+3^-4^-$ when assayed serologically at high levels such as on L cells or leukemia cells. At the low level of expression on thymocytes, this molecule appears $TL.1^-2^+3^-4^-$. The $T3^c$ gene is expressed only in Tla^c leukemias (or in a small cell subpopulation that frequently becomes leukemic) and encodes a product that is $TL.1^+2^+3^-4^+$. The two leukemia cell phenotypes are explainable by either the expression of $T13^c$ alone ($TL.1^+2^+3^-4^-$) or both $T13^c$ and $T3^c$ ($TL.1^+2^+3^-4^+$). In order to test this hypothesis, one could examine $TL.1^+2^+$ leukemia cell lines to see whether they express only the 47 kDa polypeptide and not the second 50 kDa species. The TL^- phenotype of Tla^b thymocytes is probably due to the lack of the structural gene (T13) that encodes the thymocyte TL antigen.

The significance of TL expression, that is, its function in the normal and leukemic states, awaits further experimentation.

Prospects for the Future. *Tla*-specific nucleic acid probes are now available, and this should enable the TL-encoding genes from other mouse strains to be cloned and for efforts to be focused on the transcriptionally active *Tla*

region genes. It should be possible to produce reagents (oligonucleotides or S1 mapping probes) that can distinguish between the $T3^C$ and $T13^C$ genes to decide which are transcribed in leukemia cells and thymocytes, and to, for example, test the hypotheses raised in the previous section. As other thymus-specific genes are cloned, one may be able to identify sequences important in conferring tissue specificity. Having cloned *Tla* genes and being able to make hybrid *H-2/Tla* genes should make it possible to map antigenic determinants, to map structures determining biochemical behavior (like antigenic modulation), and to map functional parts of the molecule once functional questions are addressed. The availability of cloned genes encoding TL antigens should enable one to manipulate the expresson of TL antigen, introduce it into an organism in biologically unfamiliar contexts, and perhaps uncover the function of this and the other class I differentiation antigens.

References

1. Pontarotti, P. A., H. Mashimo, R. A. Zeff, D. A. Fisher, L. Hood, A. Mellor, R. A. Flavell, and S. G. Nathenson. 1985. Conservation and diversity in the class I genes of the major histocompatibility complex: Sequence analysis of a Tla^b gene and comparison with a Tla^c gene. *Proc. Natl. Acad. Sci. USA*, submitted.

2. 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 major histocompatiblity complex by DNA-mediated gene transfer. *Nature* **300**:231.

3. Flaherty, L. 1980. The Tla region antigens. *In* Role of the Major Histocompatibility Comlex in Immunology, M. E. Dorf, editor. Garland Press, New York. 33-57.

4. Michaelson, J., E. A. Boyse, M. Chorney, L. Flaherty, E. Fleissner, U. Hämmerling, C. Reinisch, R. Rosenson, and F.-W. Shen. 1983. The biochemical genetics of the Qa-Tla region. *Trans. Proc.* **15**:2033.

5. Chorney, M. J., F.-W. Shen, J.-S.- Tung, and E. A. Boyse. 1985. Additional products of the Tla locus of the mouse. *Proc. Natl. Acad. Sci. USA*, submitted.