Studies of Class I Genes in the Major Histocompatibility Complex of the BALB/c Mouse

Thesis by

Beverly Taylor Sher

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1984

(submitted September 27, 1984)

This thesis is dedicated to Ray D. Owen

in gratitude for his advice and encouragement.

Acknowledgements

Many people helped to make my Caltech experience both intellectually and personally rewarding. I would like to thank:

Lee Hood, my thesis adviser, for creating a stimulating environment in which to work;

Norman Davidson, Ed Lewis, Elliot Meyerowitz, and Ray Owen, for serving on my thesis committee;

Tommy Douglas, Carol Sibley, and Ellen Rothenberg, for instructing me in the gentle art of teaching;

Martha Bond, Kathryn Calame, Richard Douglas, Kurt Eakle, John Frelinger, Bob Goodenow, Steve Hunt, Donna Livant, Minnie McMillan, Kevin Moore, Sue Schiffer, Jim Schilling, Iwona Stroynowski, and Hoodlums too numerous to mention, for their help and friendship;

Edwin H. McConkey, for introducing me to biological research;

Bertha Jones, Maria De Bruyn, Jessie Smith, Connie Katz, Patty Bateman, Isabella Lubomirski, and Bernita Lewis, for keeping things working;

the NSF, NIH, and Caltech, for financial support;

and finally, my parents, Phyllis and Richard Taylor; my in-laws, Ralph and Judy Sher; and my husband, Marc Sher, for making it possible.

Abstract

This thesis contains the results of investigations into the structure and organization of Class I genes in the major histocompatibility complex of the BALB/c mouse.

In the body of the thesis, the sequence of the BALB/c H-2D^d transplantation antigen gene is presented. This is the first complete sequence of an H-2D^d gene and is the only genomic sequence to be in full agreement with the available protein sequence. The H-2D^d gene sequence has been used to predict the protein sequence of the H-2D^d molecule, which has been compared to the protein sequences of other Class I molecules. The H-2D^d protein sequence is no more related to that of its closely linked partner, H-2L^d, than it is to the sequence of its presumptive allele, H-2D^b, or to the sequence of the H-2K^b molecule, which is from not only another H-2 haplotype but another genetic subregion. The sequence differences between these transplantation antigens are spread throughout the molecules in a mosaic pattern that may have arisen,in part, from small gene conversion events. No obvious evidence of any recent gene conversion event affecting the H-2D^d gene was observed, however.

Three Class I genes have been cloned and mapped to the H-2D subregion in BALB/c. These include gene 16.1, whose product has not been identified; the H-2D^d gene; and the H-2L^d gene. There is serological evidence for the existence of additional H-2D-encoded transplantation antigen molecules in BALB/c, but no genes encoding these products have been identified. The sequence of

the H-2D^d gene contains potential alternative splice sites in and around the exon encoding the first external domain. Use of these splice sites could generate a transplantation antigen molecule with different serological determinants, and might help to resolve the discrepancy between the number of H-2D-subregion Class I genes and the number of serologically defined H-2D-subregion transplantation antigens.

The appendices contain the results of a number of studies related to Class I gene organization and function. Appendix A contains the sequence of gene 27.1. This gene, also known as the Q8 gene, was identified as a Qa pseudogene based on the presence of termination codons in inappropriate locations in its sequence. Appendix B contains the sequence of the H-2L^d gene, which was the first transplantation antigen gene to be sequenced. Appendix C contains the results of DNA-mediated gene transfer experiments that identified genomic clones containing the H-2Kd, H-2Ld, and H-2Dd transplantation antigen genes as well as Class I genes encoding the Qa2,3 molecule and two different TL differentiation antigen genes. Appendix D contains the results of calculations of protein sequence homology between different Class I molecules.

Table of Contents

Figures and Figure Legends

Chapter I

Introduction

The major histocompatibility complex (MHC) is made up of a group of tightly linked loci whose products are involved in the immune response (1-4). Such genetic regions have been identified in all vertebrates examined so far. It has even been suggested that the colonial tunicate, Botryllus, a protochordate, may possess a genetic region in which loci with MHC-like functions are linked (3). Thus, the association of loci whose products are involved in immune processes is evolutionarily ancient.

The major histocompatibility complex of the mouse was the first MHC to be discovered and has been extensively characterized (1,2,4). It has been divided into three major genetic regions: the H-2 complex, the Qa region, and the Tla complex, as diagrammed in Figure 1. The H-2 complex itself has four subregions: H-2k, H-2I, H-2S, and H-2D. Each genetic region contains several loci, many of which are polymorphic. Inbred strains of mice possess different sets of MHC alleles. These constellations of alleles are referred to as haplotypes.

Three major classes of proteins are encoded by genes in the MHC: Class I, Class II, and Class III (1,2,4). Class I proteins can be separated into two categories: the Qa and TL differentiation antigens are encoded by genes in the Qa and Tla regions, while the transplantation antigens are encoded by genes in the H-2K and H-2D subregions. The Ia molecules, also known as the immune response-associated antigens, are Class II molecules encoded by genes in the H-2I subregion. Class III proteins include several complement components as well as other serum proteins and

are encoded by genes in the H-2S subregion. As this thesis is concerned primarily with transplantation antigen genes, the Class II and Class III molecules will not be further discussed.

Class I molecules share similar structures but differ in tissue distribution and extent of serological polymorphism (1,2,4,5). These molecules are ^45,000 dalton integral membrane glycoproteins found noncovalently associated with β-2 microglobulin, a 12,000 dalton polypeptide encoded by a gene outside the MHC. The structure of a typical Class I molecule is shown in Figure 2. They are divided into three external domains, each of ^90 amino acids; a hydrophobic transmembrane segment; and a carboxy-terminal cytoplasmic segment. The third external domain resembles both the β-2 microglobulin molecule and immunoglobulin domains, leading to the suggestion that Class I molecules, β-2 microglobulin, and immunoglobulins are evolutionarily related (1-4). The Tla and Qa differentiation antigens are only moderately polymorphic and have a limited tissue distribution, being found on the surfaces of particular hematopoietic cells. The transplantation antigens, in contrast, are found on the surfaces of practically all somatic cells and are among the most polymorphic molecules known. There are over 40 serologically detectable alleles at the H-2K locus and as many alleles at the H-2D locus (4).

The functions of the different types of Class I molecules also differ. The Qa and Tla differentiation antigens have no known function. The functions of the transplantation antigens are better understood. They are strongly antigenic and serve as

primary targets in the rejection of grafts between mice of different H-2 haplotypes. In addition, their presence is required for cytotoxic T cell recognition of viral or tumor antigens on the surfaces of virally infected or transformed cells. Cytotoxic T cells of the proper specificity from mice of a particular H-2 haplotype can kill virally infected cells only if these cells express H-2K or H-2D-encoded transplantation antigens of the same haplotype (6). Thus, cytotoxic T cells must have a dual specificity: they must be able to recognize not only a particular foreign antigen, but also self H-2K or H-2D molecules. This phenomenon is known as H-2 restriction. It is unclear whether the cytotoxic T cell has two receptors, one for self H-2 molecules and one for antigen, or whether cytotoxic T cells use a single receptor to recognize a combination of the two. Other molecules have been implicated in the interaction between cytotoxic T cells and their targets. For example, Lyt-2, a T cell-specific molecule, may help to stabilize the interaction between the T cell receptor(s), the transplantation antigen, and the foreign antigen (7).

Mice of different H-2 haplotypes generally express at least one H-2K-encoded transplantation antigen, known as $H\text{-}2K^x$, and at least one H-2D-encoded transplantation, known as $H\text{-}2D^x$, where x is the H-2 haplotype designation. For example, C57Bl/10 mice ($H\text{-}2^b$ haplotype) express two transplantation antigens, $H\text{-}2K^b$ and $H\text{-}2D^b$ (5). BALB/c mice (H-2d haplotype) express at least three transplantation antigens: $H\text{-}2K^d$, $H\text{-}2D^d$, and $H\text{-}2L^d$. $H\text{-}2K^d$ is encoded by the H-2K subregion, and $H\text{-}2D^d$ and $H\text{-}2L^d$ are encoded by genes in the H-2D subregion. All three of these transplantation antigens

have been purified and partially sequenced (5). In addition to these well-defined transplantation antigens, there may be more transplantation antigens encoded by H-2^d haplotype genes. In particular, there is serological evidence that the H-2D subregion in BALB/c mice may encode several other transplantation antigens, including H-$2M^d$ and H-$2R^d$ (8,9).

Class I cDNA probes have been used to determine the number of Class I genes in different inbred strains of mice and to isolate genomic clones containing Class I genes from recombinant libraries. Although the number of Class I genes appears to vary among different inbred strains, there are 25-40 Class I genes per inbred strain. Approximately 30 distinct Class I genes have been isolated from BALB/c genomic libraries (10,11). These genes have been characterized by two different approaches: DNA-mediated gene transfer and restriction polymorphism mapping (10-13). In the first approach, a Class I gene is introduced into mouse L cells (H-2^k haplotype) by the calcium phosphate technique. The resulting transfectants are then screened by radioimmunoassay with a panel of monoclonal antibodies specific for different Class I molecules, as described in Appendix C. In the second approach, low-copy DNA probes are obtained from the sequences flanking each distinct Class I gene. These probes are then used in Southern blots of genomic DNA from various inbred, congenic, and recombinant congenic strains of mice. By correlating the observed bands with the known recombination points in these strains, it is possible to map a particular probe to a locus within the MHC. This presumably also locates the Class I gene

associated with the probe sequence. In BALB/c, using these approaches, two Class I genes have been mapped to the H-2K subregion; three genes have been mapped to the H-2D subregion, as discussed in Chapter 2; and the rest of the cloned Class I genes have been mapped to the Qa and Tla regions (13).

In contrast to the situation in the H-2^d haplotype, only one H-$2D^b$ haplotype Class I gene mapping to the H-2D subregion has been cloned (14). This gene encodes the H-$2D^b$ protein. Comparisons of the protein sequences of the H-2D- encoded transplantation antigens from the H-2^d and H-2^b haplotypes have brought out some interesting points. Most transplantation antigens are only 80%-85% homologous to each other at the protein level (5). This is true even for the products of allelic genes, such as H-$2K^d$ and H-$2K^b$. This was also true in comparisons between the 183 residues of partial H-$2D^d$ protein sequence and the H-$2L^d$ and H-$2D^b$ protein sequences (15). Surprisingly, the protein sequences of H-$2L^d$ and H-$2D^b$ are 96% homologous to each other (15). This unprecedented level of homology has led to the suggestion that the H-$2L^d$ and H-$2D^b$ genes must be allelic, and that the H-$2D^d$ gene must be derived from a different locus. To elucidate the relationship between the H-$2D^d$ protein and the other H-2D-encoded transplantation antigens, it is necessary to know the sequence of the H-$2D^d$ gene. This sequence and its relationship to the other H-2D-region transplantation antigen gene sequences are presented in Chapter II.

Figure 1. The major histocompatibility complex of the mouse is diagrammed in this figure. The first line indicates that this complex maps to chromosome 17. The next line indicates protein-coding loci; the class of protein produced by each locus is indicated on the third line; the fourth line indicates the boundaries of the H-2 and Tla complexes; the fifth line shows the regions into which these complexes are divided; and the last line gives the genetic distances between these regions.

Figure 2. This figure shows the domain structure of a Class I molecule. $\alpha 1$, $\alpha 2$, and $\alpha 3$ are the three external domains of the protein. $\alpha 1$ contains the amino terminus. $\alpha 3$ associates with β-2 microglobulin, which is designated β-2m in the figure. The cytoplasmic tail of the molecule contains the carboxyl terminus.

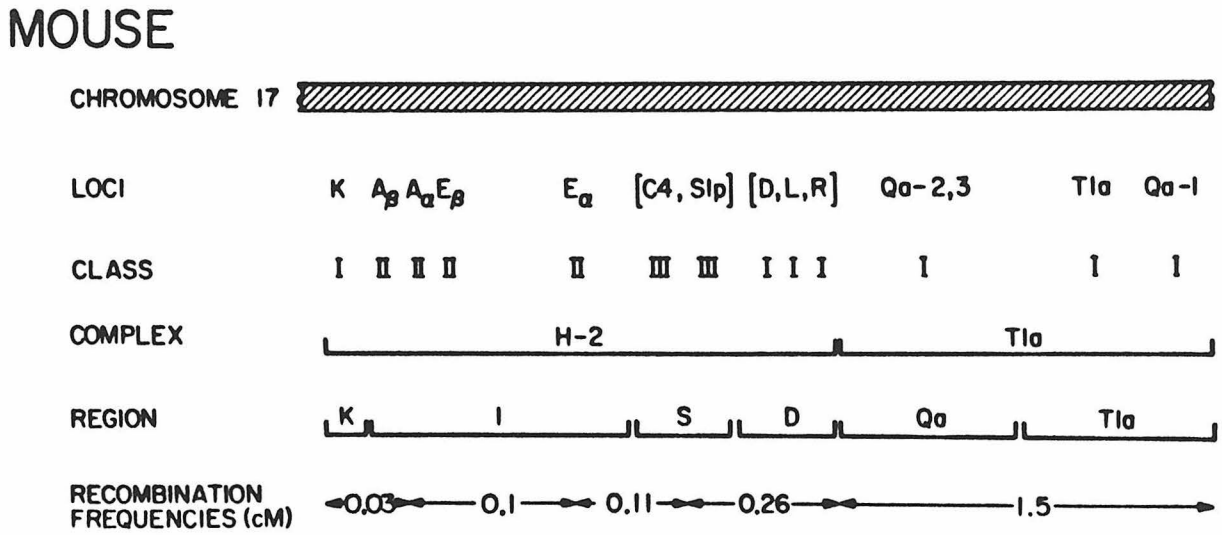

MOUSE
CHROMOSOME 17
LOCI
K
A_β A_α E_β
E_α
[C4, Slp]
[D,L,R]
Qa-2,3
Tla
Qa-1
CLASS
I II II II
II
III III
I I I
I
I
I
COMPLEX
H-2
Tla
REGION
K
I
S
D
Qa
Tla
RECOMBINATION FREQUENCIES (cM)
0.03
0.1
0.11
0.26
1.5

Figure 1.

CLASS I

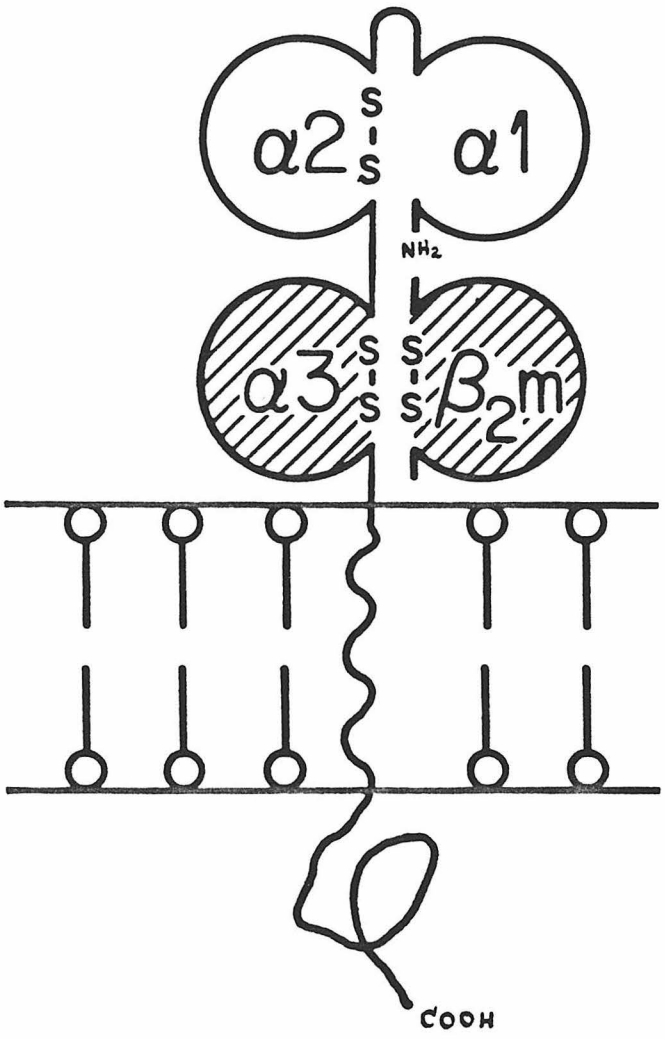


Figure 2.

References

1. Snell, G.D. (1981) Science **213**, 172-178.
2. Hood, L., Steinmetz, M., & Malissen, B. (1983) Ann. Rev. Immunol. **1**, 529-568.
3. Scofield, V. L., Schlumpberger, J. M., & Weissman, I. L. (1982) Amer. Zool. **22**, 783-794.
4. Klein, J. (1979) Science **203**, 516-521.
5. Nathenson, S. G., Uehara, H., Ewenstein, B. M., Kindt, T. J., & Coligan, J. E. (1981) Ann. Rev. Biochem. **50**, 1025-1052.
6. Zinkernagel, R. M., & Doherty, P. C. (1980) Adv. Immunol. **27**, 51-177.
7. Swain, S. L. (1983) Immun. Rev. **74**, 129-142.
8. Ivanyi, D., & Demant, P. (1979) Immunogenetics **8**,539-550.
9. Hansen, T. H., Ozato, K., Melino, M. R., Coligan, J. E., Kindt, T. J., Jandinski, J. J., & Sachs, D. H. (1981) J. Immunol. **126**, 1723-1716.
10. Steinmetz, M., Winoto, A., Minard, K., & Hood, L. (1982) Cell **28**, 489-498.
11. Hunt, S., Fisher, D., personal communication.
12. Goodenow, R. S., McMillan, M., Nicolson, M., Sher, B. T., Eakle, K. A., Davidson, N., & Hood, L. (1982) Nature **300**, 231-237.
13. Winoto, A., Steinmetz, M., & Hood, L. (1983) Proc. Natl. Acad. Sci. USA **80**, 3425-3429.
14. Weiss, E. H., Golden, L., Fahrner, K., Mellor, A., Devlin, J. J., Bullman, H., Tiddens, H., Bud, H., & Flavell, R. A. Nature, in press.

Chapter II.

DNA Sequence of the Mouse H-2D^{d} Transplantation Antigen Gene

This manuscript will be submitted to The Proceedings of the National Academy of Sciences.

Classification: Genetics

DNA sequence of the mouse H-2D^d transplantation antigen gene

(transplantation antigens/Class I genes/H-2D^d protein sequence)

BEVERLY TAYLOR SHER, RODERICK NAIRN*, JOHN E. COLIGAN‡, and LEROY E. HOOD

Division of Biology, California Institute of Technology, Pasadena, CA 91125

*Department of Microbiology and Immunology, University of Michigan Medical School, 6643 Medical Science Building II, Ann Arbor, Michigan 48109.

‡Laboratory of Immunogenetics, Building 5, Room B1-04, NIAID, NIH, Bethesda, Maryland 20205.

ABSTRACT The inbred BALB/c mouse has three transplantation antigens, H2-K^d, H2-L^d, and H2-D^d. We present the complete nucleotide sequence of the H2-D^d gene as well as 77 residues of previously unpublished H-2D^d protein sequence. These data complete the sequences of all the BALB/c transplantation antigen genes and permit detailed comparison with each other and with their counterparts from the inbred C57Bl/10 mouse. Transplantation antigens may differ from one another by as much as 5-15% of their amino acid sequence for the external domains. These extensive differences may arise by gene conversion.

The H-2D region of the BALB/c mouse encodes the H2-D^d and the H2-L^d genes. Serologic data suggests that at least two additional transplantation antigen molecules, H2-R^d and H2-M^d, are encoded in the H-2D region of the major compatibility complex. Paradoxically, gene cloning studies have only identified the H2-D^d and the H2-L^d genes in the H-2D region. A complete DNA sequence of the H2-D^d gene shows that a variety of alternative splice sites exist throughout the gene which may lead to additional gene products and may explain the multiplicity of H-2D-encoded polypeptides.

The major histocompatibility complex of the mouse (MHC), located on chromosome 17, has four regions which contain class I genes: H-2D, H-2K, Qa, and the Tla complex (1-4). Class I molecules are heterodimers consisting of a 45,000 dalton integral membrane glycoprotein noncovalently associated with β_2-microglobulin, a 12,000 dalton polypeptide encoded by a gene on chromosome 2. These molecules can be divided into two categories based on differences in their expression, the extent of their serologically detectable polymorphism, and their functions. The class I heavy chains encoded by the Tla complex are differentiation antigens of unknown function which are expressed on the surfaces of certain hematopoietic cells, and are only moderately polymorphic. The transplantation antigens, encoded by genes in the H-2D and H-2K regions, are among the most polymorphic molecules known, as over 50 different alleles have been described for each of the H-2D and H-2K loci so far (3). Different strains of mice possess different sets of alleles at MHC loci. These constellations of alleles are known as haplotypes. Transplantation antigens are found on the surfaces of virtually all somatic cells. Cytotoxic T cells recognize cell-surface viral or tumor antigens only in the context of a self transplantation antigen, a phenomenon known as H-2 restriction (5). Transplantation antigens are therefore also known as restriction elements.

Recently, a number of transplantation antigen genes from several different haplotypes have been isolated and characterized (6-14). Two different transplantation antigen genes, the H-2K^b and H-2D^b genes, have been isolated from C57Bl/10 mice (H-2^b haplotype), and three

transplantation antigen genes, the H-2K^d, H-2L^d, and H-2D^d genes, have been isolated from BALB/c mice (H-2^d haplotype). The H-2K genes map to the H-2K region of the MHC, while the H-2D and H-2L genes map to the H-2D region (15). In BALB/c mice, there is serological evidence that the H-2D region may encode other transplantation antigens, including H-2M^d and H-2R^d (16,17). Genes encoding these proteins have not yet been isolated.

The H-2D-region transplantation antigens H-2D^d, H-2L^d, and H-2D^b are particularly interesting because of their structural interrelationships. Most transplantation antigens are only 80-85% homologous to each other at the protein sequence level(1-4). This is true for presumptive allelic gene products such as H-2K^b and H-2K^d. This is also true in comparisons between the 183 residues of partial H-2D^d protein sequence and the translated coding sequences of the H-2D^b and H-2L^d genes (9,11,18,19). Surprisingly, the protein sequences of H-2L^d and H-2D^b are 95% homologous to each other. This unprecedented level of homology has led to the suggestion that the H-2L^d and H-2D^b genes are allelic (14,19), and that H-2D^d gene may be derived from a different locus. However, only one H-2D subregion H-2^b haplotype Class I gene has been cloned, namely, the H-2D^b gene (8). Thus, the number of Class I genes in the H-2D subregion of different haptotypes is variable, and the allelic relationships between the H-2D subregion Class I genes of different haplotypes is unclear. In order to shed light on the relationship between H-2D^d and the other H-2D-encoded transplantation antigens, it is necessary to know the structure of the H-2D^d

gene. In this publication, we present substantial new H-2D^d protein sequence and the complete DNA sequence of the H-2D^d gene.

MATERIALS AND METHODS

Cosmid clone c49.2 was isolated from a BALB/c Cum sperm DNA cosmid library as previously described (12,20). The construction of M13 mp8 subclones and dideoxy sequencing were performed as described (9,21,22). Protein sequencing was performed as described (23,24).

RESULTS

The Class I Gene In Cosmid Clone c49.2 Encodes H-2D^d. Gene 49.2, whose structure and sequence are shown in Figs. 1 and 2, has been shown to encode H-2D^d on the basis of several criteria. Previously, it has been shown that mouse L cells (H-2^k haplotype) transfected with c49.2 express a new protein which reacts with the H-2D^d-specific monoclonal antibodies 34-5-8 and 34-2-12 (25). Comparison of the translated coding sequence of gene c49.2 with the available protein sequence also supports the conclusion that this is an H-2D^d gene.

The partial protein sequence of the H-2D^d molecule is presented in Fig. 3. This sequence includes 77 previously unpublished residues, eight of which are still tentatively identified. In addition, 10 residues which had previously been published as tentative are now confirmed. The translated gene c49.2 sequence is in perfect agreement with all of the available H-2D^d protein data (254 of 254 residues).

The intron/exon structure of the c49.2 H-2D^d gene presented in Figs. 1 and 2 was proposed on the basis of comparison with the available H-2D^d protein sequence and with the DNA sequences of two cDNAs, pH-2I (26) and pH-2^d-1 (27). Each cDNA is >99% homologous to the coding sequence of the H-2D^d gene. Therefore they were probably derived from H-2^d gene transcripts. The H-2D^d gene has the same general exon/intron structure as the other murine class I genes which have been characterized (6-14,28). The first exon encodes a hydrophobic signal peptide which is not present in the mature protein. The second, third and fourth exons encode the three external domains of the H-2D^d protein, α1, α2 and α3. The third and fourth exons are separated by a large intron. The fifth exon encodes the transmembrane segment of the H-2D^d protein, and exons 6, 7 and 8 encode the cytoplasmic portion of the molecule. The polyadenylation signal sequence AATAAA appears 463 nucleotides 3' to the termination codon in exon 8 (29). Thus, as has been observed for other transplantation antigen genes, the intron/exon structure of the H-2D^d gene corresponds precisely with the domain structure of the H-2D^d protein.

The H-2D^d Gene Contains Potential Alternative Splicing Signals. Recently, an alternative intron/exon organization affecting the second exon has been proposed for the H-2K^d gene (30). In a cDNA clone apparently derived from a processed H-2K^d transcript, an alternative splice acceptor site in the first intron 50 nucleotides 5' to the usual acceptor site is used. In addition, an extra intron is created by the splicing out of a region inside exon 2 containing amino acids 6-38. In the H-2D^d gene

sequence, there is no potential alternative splice acceptor site in the position corresponding to that of the alternative acceptor site in the first intron of the H-2K^d gene. There is, however, a potential alternative splice acceptor site in the first intron in frame 51 nucleotides 5' to the usual splice acceptor site for exon 2. There are also potential alternative splicing signals at amino acids 6 and 38 in the H-2D^d gene sequence. It is worth noting that use of the alternative splicing signals would remove the only protein sequence that is conserved between all Class I and Class II β chain sequences, namely the sequence VRFDSD at amino acid positions 34-39 (31). It has been suggested that this sequence could be involved in interactions with T cell molecules. The potential consequences of its removal are, however, unclear.

A second region of the H-2D^d gene, the seventh intron, also contains potential alternative splice sites. The H-2D^d cDNA clones pH-2^d-1 and pH-2I both contain eighth exon sequences that are only five nucleotides long and correspond to the exon 8 sequence indicated in Fig. 2. There is a potential alternative splice acceptor sequence located 27 nucleotides 5' to the splice acceptor site used by the H-2D^d cDNAs, as indicated in Fig. 2. This potential splice acceptor site corresponds to the splice acceptor site used by the H-2K^k, H-2K^b, and H-2K^d messenger RNAs (6,7,32,33).

The presence of potential alternative splice signals in the H-2D^d gene raises the obvious question of whether they might be used. They are not used by any H-2D^d cDNAs so far described. It is interesting to note that all of the potential alternative splice acceptor sites noted in the

H-2D^d gene share the following structure: AG (C/T). All but one of the splice acceptor sites known to be used by H-2D^d transcripts have the structure AG (A/G). The only exception is the acceptor site for exon 8, which has the sequence AG T. Perhaps the AG (C/T) structure is less efficiently used as a splice site by the splicing machinery of the cell than the AG (A/G) structure.

From comparisons between the fourth exons of the H-2D^d, H-2L^d, H-2K^d, and H-2K^b genes, it is apparent that the frequency of use of a particular splice signal may depend not only on its sequence but also on its surroundings. Exon 4 of the H-2D^d gene encodes 95 amino acids, while the fourth exons of the H-2L^d, H-2K^b, and H-2K^d genes encode only 92 amino acids. This results from the absence of a splice donor site in the H-2D^d gene at a position corresponding to the position of the splice donor sites at the ends of the fourth exons of these other genes. The H-2L^d, H-2K^d, and H-2K^b genes all have a potential splice donor site which corresponds to the site at the end of the fourth exon of the H-2D^d gene, but apparently these potential alternative splice sites are not used. It is worth noting that H-2K^d messages use the rather unusual splice donor site A GT, located after the 92nd codon in exon 4, rather than splicing at the more common splice donor site G GT which is found nine nucleotides downstream. Thus, in this case, the position of the splice site, rather than its sequence, seems to dictate whether it is used.

There is a third cDNA whose DNA sequence is >99% homologous to the coding sequence of the H-2D^d gene. This is pAG64, a cDNA derived

from SV40-transformed fibroblasts (34). According to the report, this cDNA represents a transcript that is present only in SV40-transformed cells. Originally, it was not identified as an H-2D^d transcript because its sequence was not identical to that of gene Ch4a-D^d (14). Gene Ch4a-D^d, which has been partially sequenced, was identified as an H-2D^d gene, but disagrees with the H-2D^d protein sequence at 11 positions (see Discussion). The pAG64 sequence differs from the coding sequence of gene c49.2 in only 1/1412 nucleotides. Thus, it is almost certainly derived from an H-2D^d gene transcript. This cDNA is unusual in two respects. It completely lacks exon 7 sequences and it fails to terminate at the polyadenylation site used by pH-2^d-1, continuing instead for an extra 150 nucleotides. Other cDNAs derived from the same cDNA library as pAG64, pAG85 and pAG86, are identical in sequence to pAG64 except for the fact that they contain exon 7 sequences (34). Thus, it appears that exon 7 can also be involved in alternative splicing of H-2D^d transcripts.

The H-2D^d Gene Is Homologous to the Other H-2^d Haplotype Transplantation Antigen Genes. Comparisons between the sequence of the H-2D^d gene and the sequences of the other H-2^d haplotype genes, the H-2K^d and H-2L^d genes, reveal several interesting points. First, the introns are at least as related to each other as are the exons. The introns of H-2D^d gene are 87-97% homologous to the corresponding intron of the H-2L^d and H-2K^d genes, whereas the exons are 85-100% homologous at the DNA level. This differs from the situation in other multigene families, such as the globins, where the introns are much less homologous to each

other than are the exons (35, 36). Perhaps the conservation of the introns arises from gene conversion events among these genes, as will be further discussed. The only significant region of nonhomology between the DNA sequences of the H-2D^d, H-2L^d, and H-2K^d introns is associated with sequences in the third intron which are homologous to the consensus sequence of the B1 highly repetitive sequence family of the mouse (37). The areas of nonhomology result from differences in sequence and in length of the T-rich region which is located next to the B1 sequences. The B1 sequences themselves are located in homologous positions in all three genes, approximately 600 nucleotides 3' to the end of exon 3.

From the DNA sequence comparisons, it is also apparent that, for most of the length of the H-2D^d gene, it is not significantly more related to the H-2L^d gene than to the H-2K^d gene, despite the fact that the H-2D^d and H-2L^d genes are tightly linked and separated from the H-2K^d gene by approximately 0.3 centimorgans(1-3). The only region of the H-2D^d gene to display significantly higher homology to the H-2L^d gene than to the H-2K^d gene is the 3' untranslated region. Up to a point 309 nucleotides 3' to the termination codon, the 3' untranslated region of the H-2D^d gene is, if anything, more homologous to the H-2K^d gene than to the H-2L^d gene (94% versus 88% homology). After this point, however, the H-2D^d and H-2L^d genes, which remain homologous to each other and to the H-2D^b cDNA (18) diverge completely from the sequence of the H-2K^d, H-2K^b, Q10 (38), and Q8 (28) genes. Q10 is a nonpolymorphic Qa gene; Q8 is also known as the 27.1 pseudogene, and is the BALB/c equivalent of the H-2^b haplotype Q8

gene (8). As was noted in the discussion of the cDNA clone pAG64 (34), the divergence between these two groups of sequences results from the insertion of a sequence which is 84% homologous to the consensus sequence of the B2 repeated sequence family (39). This sequence is present in the 3' untranslated region of the H-2D subregion genes, but not in the other Class I 3' flanking sequences. It is flanked by nine base pair direct repeats, as indicated in Fig. 2. As direct repeats are associated with the insertion of a transposable element, it has been suggested that this sequence represents the result of a transposition of a member of the B2 family (34). This possibility is particularly attractive because the homology to the H-2K and Qa genes resumes after the second direct repeat flanking the B2 sequence.

The H-2D^d Protein Sequence Is Homologous to the Protein Sequences of Other Class I Molecules. Comparisons between the translated H-2D^d gene coding sequence and the protein sequences of other class I molecules are presented in Fig. 3. Overall, the H-2D^d protein sequence is 85% homologous to the H-2L^d protein, 84% homologous to the H-2D^b protein, 80% homologous to the H-2K^d protein, and 86% homologous to the H-2K^b protein. In general, the H-2D^d sequence is more homologous to the transplantation antigen sequences than it is to the sequences of the Qa-region molecules.

It has been suggested that the observed polymorphism of the transplantation antigens is generated, at least in part, by multiple small gene conversion events (40). Comparison of the H-2D^d gene sequence with

the sequences of the other transplantation antigen genes does not reveal any obvious signs of clearcut gene conversion events. Indeed, if the pattern of homologies observed in these comparisons was partially generated by gene conversion, it would be expected that it required many small gene conversion events that would have erased traces of previous events. In addition, even given a region in which two genes share sequence, it is difficult to determine which gene was the donor and which was the recipient in a possible gene conversion event.

DISCUSSION

The sequence of the H-2D^d gene in c49.2 is the first complete sequence of a H-2D^d gene and the only genomic sequence to be in full agreement with the available H-2D^d protein sequence data. The partial sequence of the class I gene in clone Ch4A-D^d, which was identified as a H-2D^d gene based on the reactivity of its transfected product with H-2D^d-specific monoclonal antibodies, disagrees with the H-2D^d protein sequence at 11/142 residues (14). These sequence differences are puzzling in light of the serological data, particularly as five of the changes result in charge changes. These differences could have arisen by somatic mutation, as clone Ch4A-D^d was derived from a genomic library made from MOPC41 BALB/c tumor cells. It is also possible, of course, that the Class I gene in clone Ch4a-D^d represents a second H-2D^d-like gene. If so, it is a very unusual class I gene, because class I gene introns are as homologous to each

other as class I gene exons, and introns 1 and 2 of gene Ch4a-D^d are almost impossible to align with introns 1 and 2 of our H-$2D^d$ gene. It is also possible that the differences in the Ch4A-D^d sequence arose as a result of sequencing errors.

The H-$2D^d$ protein sequence predicted from the c49.2 gene sequence is between 80-85% homologous to the protein sequences of other transplantation antigens. In particular, it is 84% homologous to the H-$2D^b$ sequence, a level of homology comparable to that found between the only previously sequenced alleles, H-$2K^b$ and H-$2K^d$ (83%). The restriction maps of the H-$2D^b$ and H-$2D^d$ genes and their flanking sequences are quite similar, as would be expected for allelic genes (8). Although it has been argued that, based on their unprecedented level of sequence homology (95%), the H-$2L^d$ and H-$2D^b$ genes are allelic (14,19), it could equally well be argued that the H-$2D^b$ and H-$2D^d$ genes are alleles, as they are as homologous to each other as are the H-$2K^b$ and H-$2K^d$ genes, and that the H-$2L^d$ gene arose from some unequal crossover event that inserted a copy of a H-$2D^b$-like gene next to the resident H-$2D^d$ gene. In a situation such as this, where there is considerable polymorphism and where there are different numbers of genes in corresponding genetic subregions, it is difficult to determine which pairs of genes are allelic.

Three distinct class I genes have been cloned and mapped to the H-$2D^d$ subregion in BALB/c mice. Several clones containing the H-$2L^d$ gene have been isolated (9-12). A second H-2D-subregion gene, located on the cosmid c16.1, has an unknown function. L cells transfected with this

gene do not produce any new cell-surface molecules that react with any of the available anti-BALB/c H-2D subregion antisera (13). Thus, this class I gene could be a pseudogene, could encode a cell-surface protein not detectable in this assay, or could encode a secreted or cytoplasmic class I-like protein. Several cosmid clones containing H-2D^d genes have been isolated (12,19,41). Their maps fall into two categories, c49.2-like and c18.1-like. The restriction maps of the two types of clones differ only in the 3' sequence flanking the H-2D^d gene. After their maps diverge, there is no DNA homology between the two types of clones. While several c49.2-like clones have been isolated, c18.1 is the only representative clone of its type (41). Based on this fact, as well as on restriction polymorphism mapping and preliminary sequence analysis, it appears that the differences in the maps are the result of a cloning artifact affecting c18.1 (42). Thus, both types of clone probably contain the same H-2D^d gene.

The presence in the H-2D^d gene of alternative splicing signals homologous to those used by other class I genes is intriguing. If used, they could generate H-2D^d-like proteins that might be useful for novel functions. Use of the potential alternative signals in and around exon 2 of the H-2D^d gene would generate cell-surface molecules that share some antigenic specificities with H-2D^d, lack other H-2D^d specificities, and have yet others of their own. This could resolve the discrepancy between the serological evidence for the existence of other H-2D-encoded molecules in the H-2^d haplotype and the lack of any additional cloned class I genes mapping to this region.

ACKNOWLEDGEMENTS

We thank Donna Livant for helping to check the H-2D^d gene sequence. We thank Marc Sher and the U.C. Irvine Proton Decay Group for computing assistance. We thank Tim Hunkapiller for useful discussions. We thank Connie Katz for typing the manuscript. We also thank Mary Spengler, Robert Valas and Michael Raum for technical assistance. The protein sequence work was supported by Grant #AI18556 to R.N. The DNA sequence work was supported by Grant #A19624 to L.H. B.T.S. was supported by N.I.H. Training Grant #GM07616.

1. Snell, G.D. (1981) *Science* **213**, 172-178.

2. Hood, L., Steinmetz, M. & Malissen, B. (1983) *Ann. Rev. Immunol.* **1**, 529-568.

3. Klein, J. (1979) *Science* **203**, 516-521.

4. Nathenson, S.G., Vehara, H., Ewenstein, B.M., Kindt, T.J., Coligan, J.E. (1981) *Ann. Rev. Biochem.* **50**:1025-1052.

5. Zinkernagel, R.M. & Doherty, P.C. (1980) *Adv. Immunol.* **27**, 51-177.

6. Kvist, S., Roberts, L. & Dobberstein, B. (1983) *EMBO J.* **2**, 245-254.

7. Weiss, E., Golden, L., Zakut, R., Mellor, A., Fahrner, K., Kvist, S. & Flavell, R.A. (1983) *EMBO J.* **12**, 453-462.

8. Weiss, E.H., Golden, L., Fahrner, K., Mellor, A., Devlin, J.J., Bullman, H., Tiddens, H., Bud, H. & Flavell, R.A. *Nature*, in press.

9. Moore, K.W., Sher, B.T., Sun, Y.H., Eakle, K. & Hood, L. (1982) *Science* **215**, 679-682.

10. Evans, G., Margulies, D.H., Camerini-Otero, R.D., Ozato, K. & Seidman, J.G. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 1994-1998.

11. Sun, Y. H., Sher, B. T., Hood, L., in preparation.

12. Steinmetz, M., Winoto, A., Minard, K. & Hood, L. (1982) *Cell* **28**, 489-498.

13. Goodenow, R.S., McMillan, M., Nicolson, M., Sher, B.T., Eakle, K., Davidson, N. & Hood, L. (1982) *Nature* **300**, 231-237.

14. Margulies, D.M., Evans, G.A., Ozato, K., Camerini-Otero, R.D., Tanaka, K., Appella, E. & Seidman, J.G. (1983) *J. Immunol.* **130**, 463-470.

15. Winoto, A., Steinmetz, M. & Hood, L. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3425-3429.

16. Ivanyi, D. & Demant, P. (1979) *Immunogenetics* **8**, 539-550.

17. Hansen, T.H., Ozato, K., Melino, M.R., Coligan, J.E., Kindt, T.J., Jandinski, J.J. & Sachs, D.H. (1981) *J. Immunol.* **126**, 1713-1716.

18. Reyes, A. A., Schold, M. & Wallace, R. B. (1982) *Immunogenetics* **16**, 1-9.

19. Reyes, A.A. & Wallace, R.B. (1982) *Immunogenetics* **16**, 265-268.

20. M. Steinmetz, unpublished data.

21. Norrander, J., Kempe, T. & Messing, J. (1983) *Gene* **26**, 101-106.

22. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. & Roe, B. A. (1980) *J. Mol. Biol.* **143**, 161-178.

23. Nairn, R., Nathenson, S.G. & Coligan, J.E. (1980) *Eur. J. Immunol.* **10**, 495-503.

24. Nairn, R., Nathenson, S.G. & Coligan, J.E. (1981) *Biochemistry* **20**, 4739-4745.

25. Goodenow, R.S., unpublished data.

26. Steinmetz, M., Frelinger, J.G., Fisher, D., Hunkapiller, T., Pereira, D. Weissman, S.M., Uehara, H., Nathenson, S.G. & Hood, L. (1981) *Cell* **24**, 125-134.

27. Bregegere, F., Abastado, J.P., Kvist, S., Rask, L., Lalanne, J., Garoff, H., Cami, B., Wiman, K., Larhammar, D., Peterson, P.A., Gachelin, G., Kourilsky, P. & Dobberstein, B. (1981) *Nature* **292**, 78-81.

28. Steinmetz, M., Moore, K.W., Frelinger, J.G., Sher, B.T., Shen, F.-W., Boyse, E.A. & Hood, L. (1981) *Cell* **25**, 683-692.

29. Proudfoot, N. (1982) *Nature* **298**, 516-517.

30. Lalanne, J.-L., Cochet, M., Kummer, A.-M., Gachelin, G. & Kourilsky, P. (1984) *Proc. Natl. Acad. Sci. USA*, in press.

31. Malissen, M., Hunkapiller, T. & Hood, L. (1983) *Science* **221**, 750-754.

32. Uehara, H., Coligan, J. & Nathenson, S.G. (1981) *Biochemistry* **20**, 5940-5945.

33. Kress, M., Glaros, D., Khoury, G. & Jay,G. (1983) *Nature* **306**, 602-604.

34. Brickell, P. M., Latchman, D. S., Murphy, D., Willison, K., Rigby, P. W. J. (1983) *Nature* **306**, 756-759.

35. van den Berg, J., van Ooyen, A., Mantei, N., Schambock, A., Grosveld, G., Flavell, R.A. & Weissman, C. (1978) *Nature* **275**, 37-44.

36. Efstratiadis, A., Posakony, J., Maniatis, T., Lawn, R.M., O'Connell, C., Spritz, R.A., De Riel, J.K., Forget, B., Weissman, S.M., Slightom, J.L., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C. & Proudfoot, N.J. (1980) *Cell* **21**, 653-668.

37. Krayev, A.S., Kramerov, D.A., Skryabin, K.G., Ryskov, A.P., Baev, A.A. & Georgiev, G.P. (1980) *Nucleic Acids Res.* **8**, 1201-1215.

38. Mellor, A.L., Weiss, E.H., Kress, M., Jay, G. & Flavell, R.A. (1984) *Cell* **36**, 139-144.

39. Krayev, A. S., Markusheva, T. V., Kramerov, D. A., Ryskov, A. P., Skryabin, K. G., Bayev, A. A. & Georgiev, G. P. (1982) *Nucleic Acids Res.* **10**, 7461-7475.

40. Brégégère, F. (1983) *Biochimie* **65**, 229-237.

41. Sun, Y.H., unpublished data.

42. Sher, B.T., unpublished data.

FIGURE LEGENDS

Figure 1. This figure shows the structure of the H-2D^d gene and sequencing strategy used to determine its sequence. A battery of restriction enzymes (HhaI, Sau3A, Sau96A, RsaI, and AluI) was used to generate shotgun mp8 subclones from the 2.2 kb Bam fragment and the 4.2 kb KpnI fragment containing the 5' and 3' ends of the gene. These subclones were sequenced by the dideoxy method. The locations of B1 and B2 sequences are indicated. The numbers given are distances in nucleotides.

Figure 2. This figure contains the sequence of the H-2D^d gene. The protein translations of the exons are given above the DNA sequence. Potential alternative splice sites are indicated by arrows.

Figure 3. The protein sequences of several class I molecules are compared. The line denoted "H-2D^d pro" gives the partial H-2D^d protein sequences. Sequence identity is indicated by a dot; sequence differences are indicated by the one-letter amino acid code. The symbols on the H-2D^d pro line are as follows: *: newly published, tentative identification; =: previously tentative, now identified; +: previously unpublished, now identified; @: previously published, still tentative; and -: previously published.

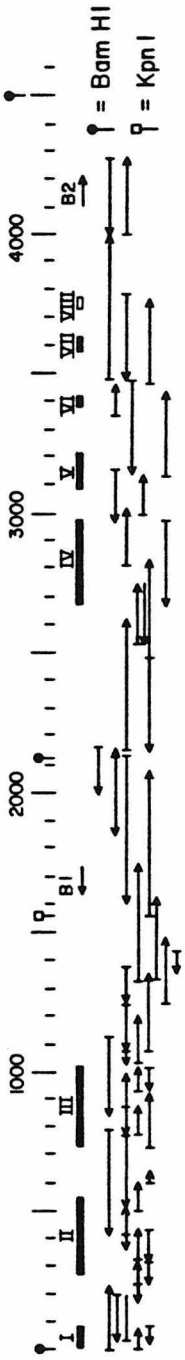


Figure 1.

Figure 2.

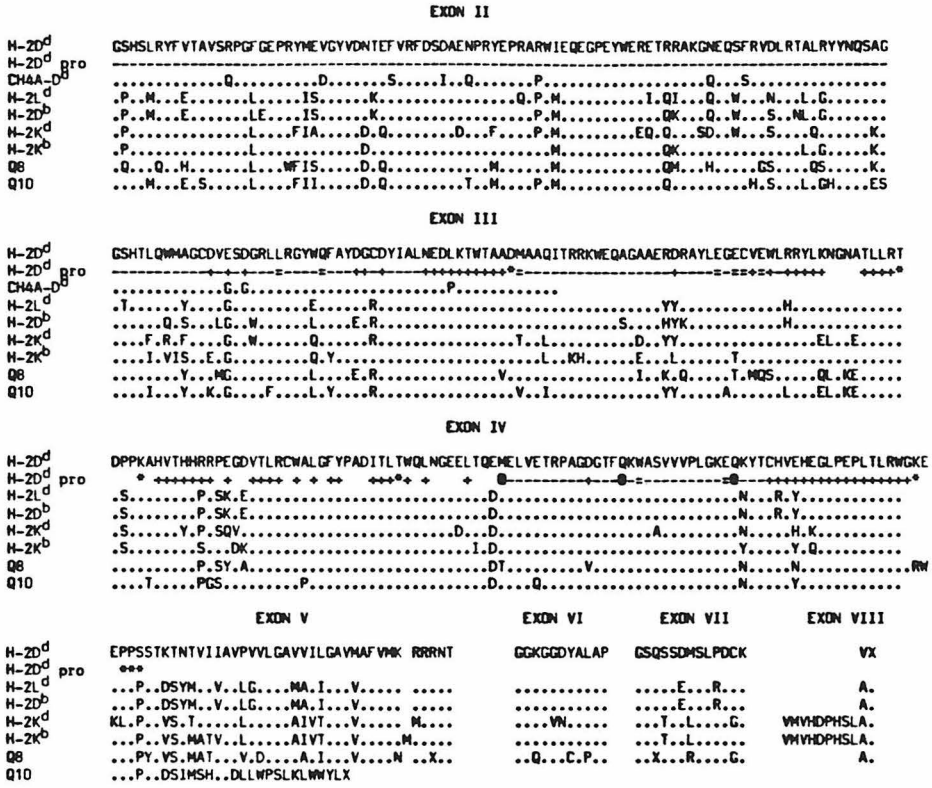

EXON II
H-2D^d GSHSLRYFVTAVSRPGFGEPRYMEVGYVDNTEFVRFDSDAENPRYEPRARWIEQEGPEYWERETRRAKGNEQSFRVDLRTALRYYNQSAG
H-2D^d pro
CH4A-D^d .....Q.....D.....S.....I..Q.....P.....Q...S.....
H-2L^d .P..M...E.....L.....IS.....K.....Q.P.M.....I.QI...Q..W...N...L.G.....
H-2D^b .P..M...E.....LE....IS.....K.....P.M.....QK...Q..W...S..NL.G.....
H-2K^d .P.....L....FIA.....D.Q.....D...F....P.M.....EQ.Q...SD..W...S....Q.....K.
H-2K^b .P.....L.....D.....M.....QK.....L.G.....K.
Q8 .Q...Q..H.....L...WFIS.....D.Q.....M.....M.....QM...H.....GS....QS.....K.
Q10 ....M...E.S.....L....FII.....D.Q.....T..M....P.M.....Q.....H.S...L.GH....ES
EXON III
H-2D^d GSHTLQWMAGCDVESDGRLLRGYWQFAYDGCDYIALNEDLKTWTAADMAAQITRRKWEQAGAAERDRAYLEGECVEWLRRYLKNGNATLLRT
H-2D^d pro
CH4A-D^d .....G.G.....P.....
H-2L^d .T.....Y....G.....E.....R.....YY.....H.....
H-2D^b .....Q.S...LG..W.....L....E.R.....S....HYK.....H.....
H-2K^d ....F.R.F....G..W.....Q.....R.....T..L.....D..YY.....EL..E.....
H-2K^b ....I.VIS..E.G.....Q.Y.....L..KH.....E...L.....T.....
Q8 .....Y...MG.....L....E.R.....V.....I..K.Q.....T.MQS.....QL.KE.....
Q10 ....I...Y..K.G....F....L.Y....R.....V..I.....YY.....A.....L...EL.KE.....
EXON IV
H-2D^d DPPKAHVTHHRRPEGDVTLRCWALGFYPADITLTWQLNGEELTQEMELVETRPAGDGTFQKWASVVVPLGKEQKYTCHVEHEGLPEPLTLRWGKE
H-2D^d pro
H-2L^d .S.....P.SK.E.....D.....N...R.Y.....
H-2D^b .S.....P.SK.E.....D.....N...R.Y.....
H-2K^d .S.....Y.P.SQV.....D...D.....A.....N.....H.K.....
H-2K^b .S.....S...DK.....I.D.....Y.....Y.Q.....
Q8 .....P.SY.A.....DT.....V.....N.....N.....RW
Q10 ....T.....PGS.....P.....D....Q.....N.....Y.....
EXON V
EXON VI
EXON VII
EXON VIII
H-2D^d EPPSSTKTNTVIIAVPVVLGAVVILGAVMAFVMK RRRNT GGKGGDYALAP GSQSSDMSLPDCK VX
H-2D^d pro
H-2L^d ...P..DSYM..V..LG....MA.I...V..... ....E...R... A.
H-2D^b ...P..DSYM..V..LG....MA.I...V..... ....E...R... A.
H-2K^d KL.P..VS.T.....L.....AIVT...V..... M.... ....VN..... ...T..L....G. VMVHDPHSLA.
H-2K^b ...P..VS.MATV..L.....AIVT...V.....M..... ....T..L..... VMVHDPHSLA.
Q8 ...PY.VS.MAT...V.D....A.I...V....N ..X.. ..Q...C.P.. ..X...R....G. A.
Q10 ...P..DSIMSH..DLLWPSLKLWWYLX

Figure 3.

Appendix A

DNA Sequence of a Class I Pseudogene

The gene whose sequence is described in this publication was the first Class I gene to be sequenced. This publication appeared in Cell.

A Pseudogene Homologous to Mouse Transplantation Antigens: Transplantation Antigens Are Encoded by Eight Exons That Correlate with Protein Domains

Cell, Vol. 25, 683–692, September 1981, Copyright © 1981 by MIT

Michael Steinmetz, Kevin W. Moore, John G. Frelinger, Beverly Taylor Sher, Fung-Win Shen,* Edward A. Boyse* and Leroy Hood
Division of Biology
California Institute of Technology
Pasadena, California 91125
*Memorial Sloan-Kettering Cancer Center
New York, New York 10021

Summary

We have isolated about 30 to 40 different BALB/c mouse sperm DNA genomic clones that hybridize to cDNA clones encoding proteins homologous to transplantation antigens. One of these clones (27.1) was selected for sequence analysis because it was polymorphic in Southern blot analyses of the DNAs from BALB/c and CBA mice. A fragment of 5.7 kilobases of this clone was completely sequenced and found to contain a pseudogene whose sequence is highly homologous to the sequences of known transplantation antigens. Pseudogene 27.1 is split into eight exons that correlate with the structurally defined protein domains of transplantation antigens. Using Southern blot hybridization on the DNAs of different inbred mouse strains, we mapped the pseudogene to the Qa-2,3 region, a part of the Tla complex on chromosome 17 that is adjacent to the major histocompatibility complex. The Qa-2,3 region encodes lymphoid differentiation antigens homologous to the transplantation antigens in size, in peptide map profiles and in their association with β2-microglobulin. These mapping studies suggest that gene 27.1 may be a pseudogene for either a Qa antigen or an as yet undefined transplantation antigen. Accordingly, we may have isolated genes encoding lymphoid differentiation antigens of the Tla complex as well as those encoding transplantation antigens among the 30 to 40 different genomic clones isolated from our sperm library.

Introduction

The major histocompatibility complex plays a fundamental role in the regulation of the vertebrate immune response. It encodes three families or classes of proteins—class I, the transplantation antigens; class II, immune-response-associated or Ia antigens; and class III, genes encoding complement components (Klein, 1975; Ploegh et al., 1981). Adjacent to the mouse major histocompatibility complex or H-2 complex on chromosome 17 is the Tla complex (Figure 1), which contains several genes encoding lymphoid differentiation antigens, the Qa and TL antigens (Flaherty, 1980).

The class I or transplantation antigens (K, D, L and R) are integral membrane proteins with molecular weights of 45,000 and are associated β2-microglobulin. The Qa and TL antigens appear to be similar in size to the class I molecules and also are associated with β2-microglobulin; this raises the possibility that the class I molecules are evolutionarily related to the Qa and TL antigens (Michaelson et al., 1977; Stanton and Hood, 1980). The transplantation antigens are expressed on most somatic cells of the mouse, whereas the TL antigens are normally expressed only on thymocytes (Boyse et al., 1964) and the Qa antigens only on hematopoietic cells (Flaherty, 1976; Stanton and Boyse, 1976; Hämmerling et al., 1979). Structural analyses of the transplantation antigens of man and mouse, including proteolytic digestions and amino acid sequence analyses, suggest that they are divided into a series of discrete domains—three extracellular or external domains, each of about 90 residues, and a transmembrane domain and a cytoplasmic domain, each of about 30 residues (Lopez de Castro et al., 1979; Coligan et al., 1981). Both the second and third external domains have a centrally placed disulfide bridge spanning about 60 residues. Thus they are similar in general structure to immunoglobulin domains, which also have a centrally located disulfide bridge. Moreover, modest amino acid sequence homology (~35%) has been noted between the third external domain of transplantation antigens and immunoglobulin domains (Orr et al., 1979; Strominger et al., 1980). These observations suggest that transplantation antigens are evolutionarily related to immunoglobulins.

A number of investigators have reported the isolation and characterization of cDNA and genomic clones encoding proteins whose sequences are highly homologous to those of known human transplantation antigens (Jordan et al., 1981; Ploegh et al., 1980; Sood et al., 1981) and mouse transplantation antigens (Kvist et al., 1981; Steinmetz et al., 1981). By DNA sequence analysis of the cDNA clones, we could show that the third external domain of the transplantation antigen shows striking homology to the constant region domains of immunoglobulins (Steinmetz et al., 1981). This observation suggests that immunoglobulins and transplantation antigens diverged from a common ancestral gene. Moreover, Southern blot analysis with these cDNA probes against germline DNA revealed that there are at least 15 cross-hybridizing species of genomic DNA (Cami et al., 1981; Steinmetz et al., 1981). Thus the genes encoding transplantation antigens constitute a multigene family.

In this paper, we report the DNA sequence of a 5.7 kilobase (kb) fragment of a mouse genomic clone (27.1) that hybridizes strongly to our cDNA probes. The DNA sequence of gene 27.1 includes eight exons, seven of which are homologous to our transplantation antigen cDNA sequences and correlate with the struc-

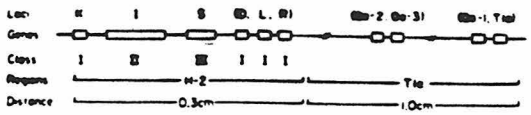


Figure 1. Genetic Map of the H-2 and Tla Complexes of Chromosome 17 in the Mouse

The order of loci within brackets is not known. Distances are in CentiMorgans.

tural domains of transplantation antigens. This gene is a pseudogene that maps in the Qa-2,3 gene cluster of the Tla complex. Hence it may be a pseudogene either for a Qa differentiation antigen or for a transplantation antigen mapping to the right of the D-marker locus of the major histocompatibility complex. Thus, using our cDNA probes, we may be able to isolate a variety of closely related genes encoding lymphoid differentiation antigens (Qa and TL) as well as those encoding the classical transplantation antigens.

Results and Discussion

Isolation of Genomic Clones Hybridizing to cDNA Probes for Transplantation Antigens

About 1.5 × 10^6 phage clones of a BALB/c sperm DNA library were screened with a mixture of our cDNA clones pH-2III, which encodes portions of the NH_2-terminal half, and pH-2I, which encodes portions of the C-terminal half of transplantation antigens (Figure 2A). Approximately 30 to 40 different clones were isolated (K. Moore, B. T. Sher and H. Sun, unpublished data). One of these genomic clones, denoted 27.1, was chosen for DNA sequence analysis because this gene exhibited polymorphic restriction sites when the DNAs from different inbred strains of mice were analyzed by Southern blot analysis (see below).

The Eight Exons of Gene 27.1 Correspond to the Structurally Defined Protein Domains of the Transplantation Antigen

Clone 27.1 was mapped with several restriction enzymes, and the coding sequences were identified by hybridization with our cDNA clones. This analysis showed that the gene present in clone 27.1 was located on a 5.7 kb Bgl II-Bam HI fragment in the middle of the 19 kb mouse DNA insert (Figure 2B). The sequence of the 5.7 kb fragment was completely determined by use of both the M13 dideoxy sequencing technique (Sanger et al., 1980; Messing et al., 1981) and the chemical degradation method (Maxam and Gilbert, 1980) (Figure 2C). The DNA sequence is given in Figure 3. A comparison of the genomic sequence with the three cDNA sequences of clones pH-2I, pH-2II and pH-2III, as well as a comparison of the translated genomic DNA sequences with the complete amino acid sequence of the K^b molecule (Coligan et al., 1981) clearly delineates exons 2–7, which correlate strikingly with the structural domains of the transplantation antigen (Figure 4). The DNA sequences of these exons in gene 27.1 are 81–95% homologous to the corresponding sequences in the three cDNA clones (Table 1). Furthermore, the translated sequences of these exons of gene 27.1 are almost as homologous to the K^b polypeptide as are the translated cDNA sequences (Table 2).

Exon 1 could not be identified by sequence comparisons with the cDNA clones and transplantation antigens. It was identified on the basis that it starts with a methionine codon, encodes a stretch of 21 amino acids, 19 of which are hydrophobic or uncharged, and contains at its 3′ end an upstream RNA splicing signal that has its downstream counterpart in the first codon of exon 2 (Figure 3). Its hydrophobicity, location and size indicate that the peptide encoded by exon 1 probably represents the signal or leader sequence identified as a component of the initial translation product of the mRNA for transplantation antigens (Dobberstein et al., 1979; Ploegh et al., 1979). Exons 2, 3 and 4 correspond to the three external domains of the transplantation antigen (Figures 3 and 4) and have as their boundaries those amino acid residues predicted from the protein sequence homology comparisons (Lopez de Castro et al., 1979; Coligan et al., 1981). The translated amino acid sequences of the extracellular domains of gene 27.1 are 73–85% homologous to the K^b molecule at the protein level (Table 2). Exon 5 corresponds to the transmembrane domain. The transmembrane domain appears to be among the least highly conserved domains (Table 2), in keeping with the functional requirement for conservation of the hydrophobic character of the transmembrane belt, but not its specific amino acid sequence. Two cytoplasmic exons, 6 and 7, can readily be identified by their homology to the cDNA sequences and the H-$2K^b$ amino acid sequence (Figures 3 and 4). Exon 8 was identified both in its coding and 3′ untranslated region by comparison with the cDNA clone pH-2I (see below for a comparison of this region with clone pH-2II). We were surprised to note that three exons, 6, 7 and 8, encode the cytoplasmic portion of the molecule. There is no evidence at the protein level to suggest that the cytoplasmic region is comprised of more than one functional domain. However, the cytoplasmic exons may reflect distinct functional entities in that there is a possibility that the 3′ end of the gene may use alternative RNA splicing patterns that lead to distinct cytoplasmic domains.

The exon encoding the third extracellular domain is highly conserved and shows a narrower range of variability than any of the other five exons (Table 1). Since only the third domain in the transplantation antigen shows significant homology to the constant region domains of immunoglobulins, such restricted divergence may reflect the constraints imposed by its interaction with yet another immunoglobulin-like do-

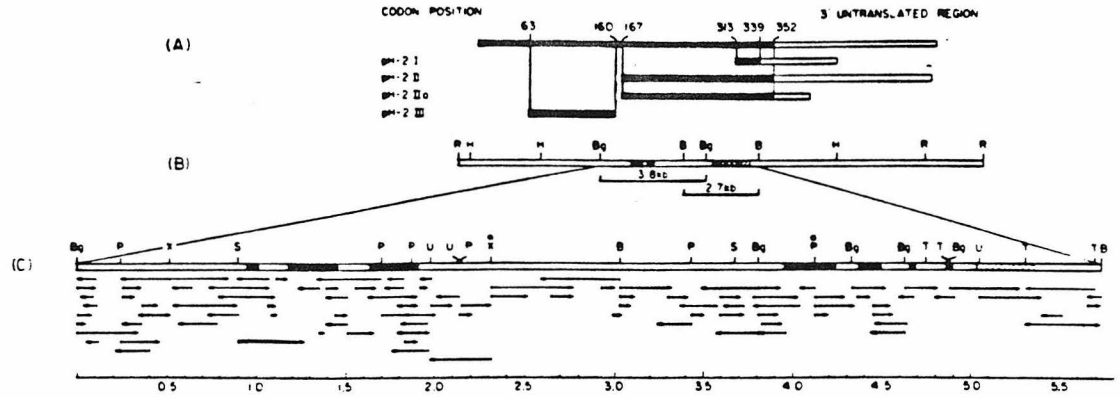


Figure 2. Maps of Three cDNA Clones and the Genomic Clone 27.1

(A) A comparison of the cDNA probes used in this paper and the coding and 3′ untranslated regions of mRNAs for transplantation antigens (see Steinmetz et al., 1981). (B) Partial restriction map of the 19 kb insert found in clone 27.1. The Eco RI sites at the ends of the insert were generated during construction of the BALB/c sperm DNA library from partial Hae III–Alu I genomic fragments and Eco RI linker fragments (Davis et al., 1980). The 3.8 kb Bgl II fragment and the 2.7 kb Bam HI fragment were isolated for DNA sequence analysis (see Experimental Procedures). The location of the exons of gene 27.1 are indicated by black boxes. A hatched box denotes the 3′ untranslated region. (B) Bam HI, (Bg) Bgl II, (H) Hind III, (R) Eco RI. (C) Map of the 5.7 kb Bgl II–Bam HI fragment of clone 27.1 showing the position of individual sequences obtained. The straight arrows give the 5′–3′ direction and the length of the sequences obtained from individual M13 clones. Wavy arrows indicate the direction and length of sequences obtained by the chemical degradation method (see Experimental Procedures). At the Xba I site and the Hae III site 4 bp upstream of the Pst I site marked by a star, no overlapping sequences were obtained. (B) Bam HI; (Bg) Bgl II; (P) Pst I; (S) Sal I; (T) Sst I; (U) Pvu II; (X) Xba I.

main, β2-microglobulin. Thus this domain may have rigid selective constraints placed on its divergence in different genes for transplantation antigens. Additional sequence data on the class I genes or proteins should reveal whether diversity is clustered in certain regions of the exons coding for extracellular domains. The 3′ untranslated region of these genes is as highly conserved as the coding regions.

Gene 27.1 Is a Pseudogene by Several Criteria

The sequence presented in Figure 3 shows that two stop codons are found in the reading frame at positions 312 in the transmembrane exon and 328 in the second cytoplasmic exon. The part of the transmembrane exon containing the termination codon has been determined by dideoxy sequencing of both strands as well as by the chemical degradation method (Figure 2C). The two stop codons would not allow the translation of an intact class I polypeptide of gene 27.1. Moreover, in the exon encoding the transmembrane domain, codon 292 codes for the charged residue, aspartic acid. The presence of an aspartic acid in this position would prevent an energetically stable integration of the transmembrane domain into the lipid bilayer. In addition, the DNA sequence at one exon–intron junction is at variance with the consensus sequences that are believed to be necessary for RNA splicing (Lewin, 1980). The acceptor site of the second cytoplasmic exon (exon 7 in Figure 3) shows the deletion of a G in the last position of the intervening sequence when compared with the consensus sequence. A shift of the splice point by one nucleotide into the exon might generate a functional splice site but would alter the reading frame for this exon.

One other feature of gene 27.1 distinguishes it from the clone pH-2II. At least 165 bp at the end of the 3′ untranslated region of gene 27.1 appear to have been inverted and translocated into the intervening sequence between exons 3 and 4 (Figure 5). Since this part of the untranslated region in pH-2II is a repetitive element in the mouse genome (Steinmetz et al., 1981), it also is possible that the sequence between exons 3 and 4 in gene 27.1 represents another copy of this repetitive sequence and that the corresponding homologous region at the 3′ end of gene 27.1 was deleted.

Although gene 27.1 is a pseudogene, it shows striking homology to the K^b polypeptide (Table 2) as well as to the three cDNA clones (Table 1). Moreover, the intron–exon boundaries of the first external domain are identical to those of one other genomic clone that appears to be expressed as a transplantation antigen in cells transformed with this clone (R. Goodenow and J. A. Frelinger, unpublished data) and is now being characterized (K. Moore and B. T. Sher, unpublished). Thus pseudogene 27.1 retains most of the fundamental characteristics of the genes encoding transplantation antigens.

Gene 27.1 Maps to the Tla Region of Mouse Chromosome 17

We have previously shown that a restriction enzyme polymorphism is observed when Bam HI–digested DNAs from BALB/c (d haplotype) and CBA (k haplotype) mice are probed with clone pH-2III in a Southern blot analysis (Steinmetz et al., 1981). A strongly hy-

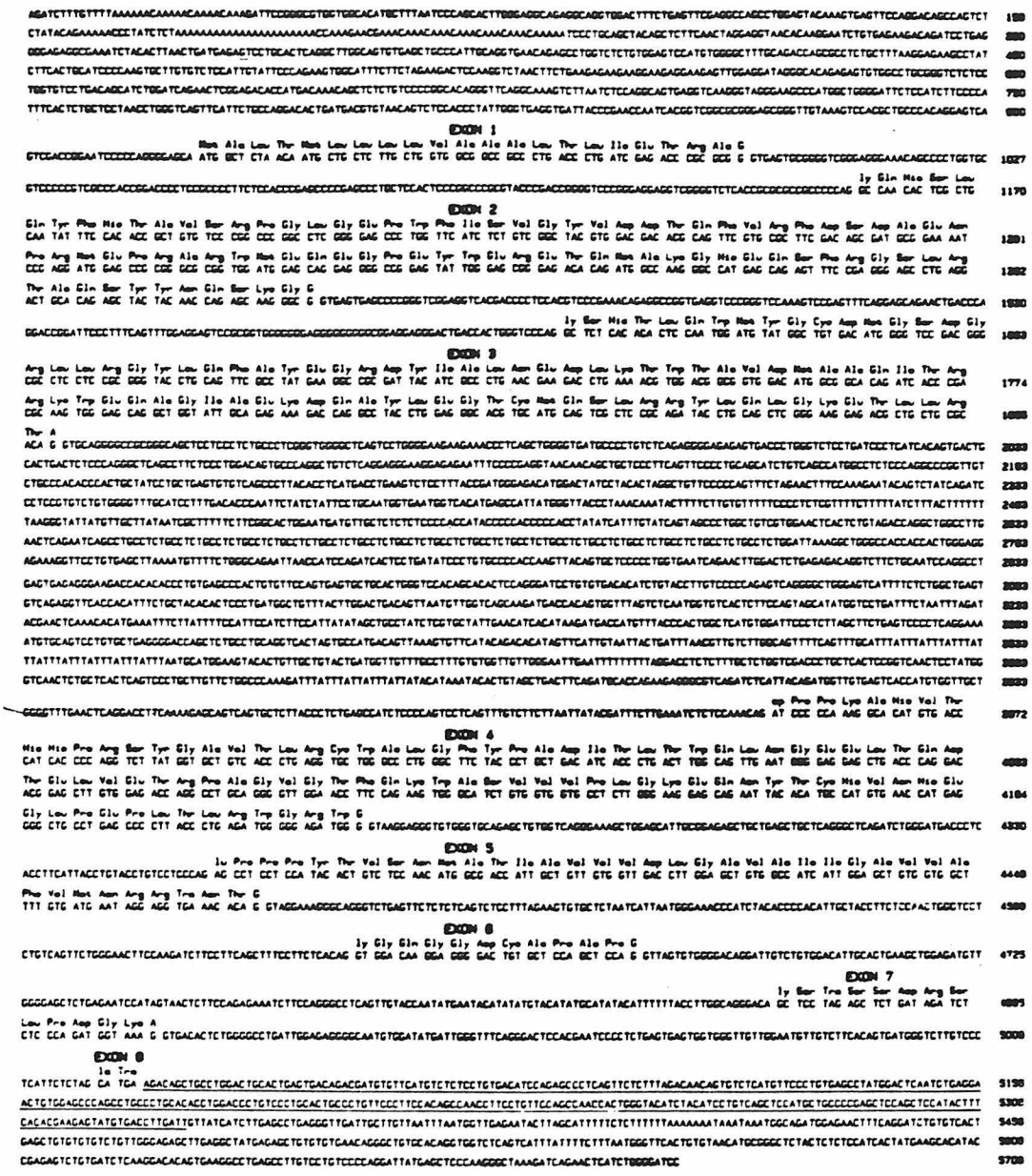


Figure 3. The DNA Sequence of the 5.7 kb Bgl II-Bam HI Fragment of Clone 27.1

Amino acid sequences encoded by the exons of gene 27.1 are shown above the DNA sequences. The 3' untranslated region is underlined.

bridizing 6.2 kb band seen in BALB/c DNA is replaced by a 5.4 kb band in CBA DNA. In a similar experiment in which pH-2IIa is used as a probe (Figure 2A), a strongly hybridizing 2.7 kb Bam HI fragment (lower band of the doublet) is missing in the k haplotype DNA (Figure 6). Clone 27.1 was initially selected from among the 30 to 40 distinct genomic clones from the BALB/c library because it contained a 6.2 kb Bam HI fragment hybridizing to clone pH-2III and a 2.7 kb Bam HI fragment hybridizing to clone pH-2IIa. This clone therefore was a candidate for the BALB/c DNA sequence showing the polymorphic restriction patterns described above. To test this possibility, we also have analyzed clone 27.1 by Sst I and Xba I digestions and compared it with similarly digested BALB/c and CBA DNAs probed with clone pH-2IIa. For both enzymes, a strongly hybridizing band in BALB/c DNA corresponded in size to that restriction fragment of

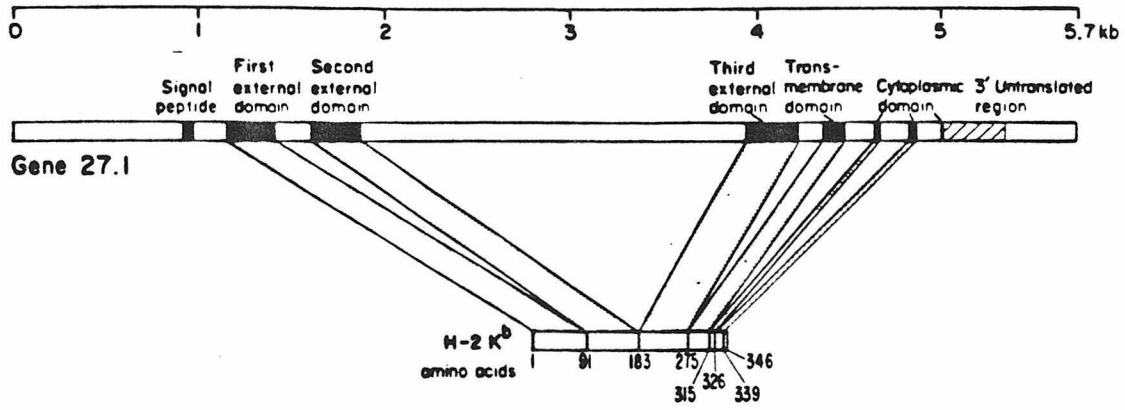


Figure 4. Organization of Exons and Introns in Gene 27.1

Gene 27.1 is compared to the H-2K^b transplantation antigen (Coligan et al., 1981) to show the location of intervening sequences with respect to the amino acid positions. Exons 1 to 8 labeled according to their corresponding protein domains are shown as black boxes and the 3' untranslated region as a hatched box. Compared to the H-2K^b molecule, the third external domain of gene 27.1 is longer by three codons at its 3' end. Three additional codons at this position have also been found in the cDNA clone pH-2^d-1 (Brégegère et al., 1981). Codon 309 (methionine in H-2K^b) is not found in gene 27.1. It is also absent in the cDNA clones pH-2II (Steinmetz et al., 1981) and pH-2^d-1 (Kvist et al., 1981). The intervening sequence between the second and the third cytoplasmic exon is postulated from a comparison between clones 27.1 and pH-2I. For a comparison of this part of the 27.1 sequence against pH-2II see Figure 9. The homology between the translated sequence of gene 27.1 and the H-2K^b protein sequence ends precisely at the beginning of this 139 bp intron. A DNA sequence that might code for a protein sequence homologous to amino acids 340–346 of H-2K^b has not been identified in gene 27.1.

Table 1. A Comparison of the DNA Sequences (Percent Homology) of the Three cDNA Clones and the Exons in Gene 27.1[a]

	pH-2I	pH-2II	pH-2III
First external domain (exon 2)			81%
Second external domain (exon 3)		84%	87%
Third external domain (exon 4)		95%	
Transmembrane domain (exon 5)		87%	
First cytoplasmic exon (exon 6)	88%	88%	
Second cytoplasmic exon (exon 7)	90%	88%[b]	
Third cytoplasmic exon and 3'-untranslated region (exon 8)	85%	84%[c]	

[a] Sequences were compared for the individual domains (Figures 3 and 4) where overlaps were present.

[b] For the comparison of the 3' end of gene 27.1 with the cDNA clone pH-2II, a different splicing pattern was assumed as depicted in Figure 9.

[c] The comparison of the 3' untranslated region between gene 27.1 and clone pH-2II was somewhat complex in that a portion of the region in gene 27.1 was translocated adjacent to the third exon (Figure 5). Homology of 87% was found for the region 3' to the second cytoplasmic exon, and homology of 80% was found for the translocated 3' region. The mean is given in the table.

Table 2. Comparison of the H-2K^b Polypeptide (Percent Homology) with the Translated Sequences of Clones pH-2I, pH-2II and pH-2III and the Exons of Gene 27.1[a]

	pH-2I	pH-2II	pH-2III	27.1
First external domain			86%	82%
Second external domain		94%	90%	73%
Third external domain		88%		85%
Transmembrane domain		67%		64%
First cytoplasmic exon	100%	100%		73%
Second cytoplasmic exon	85%	89%		62%

[a] The sequences were compared where overlapping regions were present. For the second cytoplasmic exon of the K^b molecule, only amino acids 326–338 were compared because the C-terminal seven residues are not homologous to any of the corresponding DNA sequences (text).

27.1 carrying the coding sequence. In each case, this band was missing in CBA DNA. Thus we are confident that clone 27.1 represents the identified polymorphic DNA sequence.

To map gene 27.1, we used a series of congenic and recombinant lines of mice. Congenic lines are strains of mice genetically identical to a second strain except for a limited segment of one chromosome that carries a different allele at one or more marker loci (Klein, 1975). Recombinant congenic lines in the Tla complex have been developed from recombinational events between two congenic lines (Flaherty, 1976). Any differences observed between recombinant congenic strains in the expression of a given trait suggest that the gene controlling that trait is identical or closely linked to the marker locus. When we examined by Southern blot analysis the congenic recombinant strains B6.K1 and B6.K2, which arose from recombinations between the congenic lines B6-Tla^a and B6-H-2^k (Flaherty, 1976), we found that the 2.7 kb Bam HI fragment was present in the B6.K2 strain and absent in the B6.K1 strain (Figure 6). When the appearance of this trait is compared with the genetic maps of these strains (Figure 7), this trait maps to the right of the recombinational site of the B6.K1 strain and to the left of the recombinational site of the B6.K2

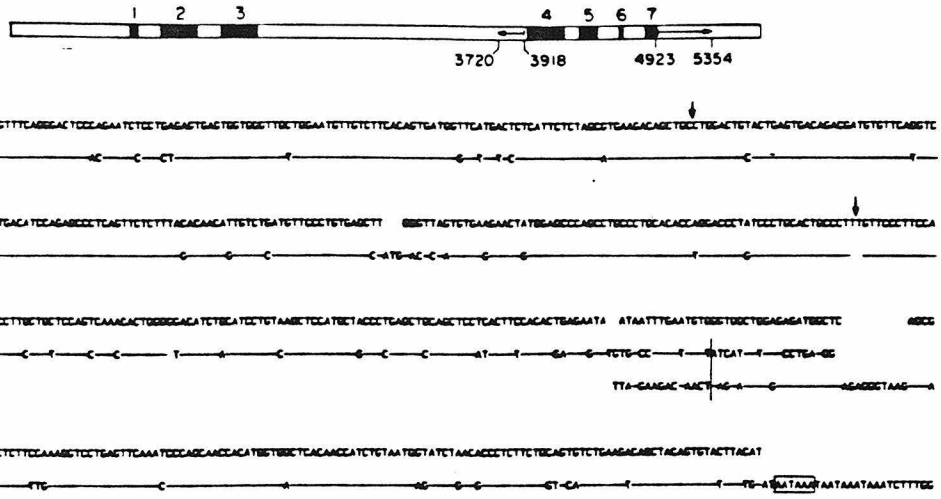


Figure 5. A Comparison of the DNA Sequences of the 3′ Untranslated Region of the cDNA Clone pH-2II and the Homologous Sequences Found in Gene 27.1

The exon–intron structure of gene 27.1 as derived from the comparison with pH-2II (see Figure 9) is shown in the top part of the figure with the arrows indicating the extent and 5′ to 3′ directions of the sequences that are compared to the cDNA sequence in the lower part of the figure. Nucleotide numbers refer to Figure 3. The stop codon at the beginning of the 3′ untranslated sequence in pH-2II is boxed. Sequence AATAAA, which is thought to be a recognition signal for poly(A) addition in most eucaryotic mRNAs (Proudfoot and Brownlee, 1976), is also indicated in gene 27.1. The DNA sequence of clone pH-2II was previously published (Steinmetz et al., 1981), except for a stretch of 172 bp (indicated by arrows) in the middle of the 3′ untranslated region. This sequence was determined by dideoxy sequence analysis of a M13mp2 phage containing the 717 bp Pst fragment covering this region (see Figure 1 in Steinmetz et al., 1981). The cloning of this fragment into the Eco RI site of M13mp2 and the DNA sequence analysis were performed as described in Experimental Procedures for fragments derived from clone 27.1.

strain—that is, in the Qa-2,3 region. The exact location of the recombination point in strain B6.K1 between the H-2D and Qa-2,3 regions is not known. Thus gene 27.1 may be either a Qa pseudogene or a pseudogene for a transplantation antigen mapping within the D region, but to the right of the D-marker locus. Soloski et al. (1981) have shown there are peptide homologies between the Qa-2 antigens and transplantation antigens. These protein homologies together with the mapping of gene 27.1 therefore raise the possibility that our cDNA probes may allow us to isolate not only genes encoding transplantation antigens but also genes encoding Qa, and perhaps other genes for lymphoid differentiation antigens encoded with the Tla complex (Figure 1). Thus the 30 to 40 genomic clones that we have isolated from the sperm library may include genes for Tla differentiation antigens as well as those for transplantation antigens.

The Intervening Sequence between the Seventh and Eighth Exons of Clone 27.1 Is Present in the cDNA Clone pH-2II

A comparison of the cDNA clones pH-2I and pH-2II shows that these sequences, though encoded by distinct genes, are highly homologous except for a 139 bp insertion found in clone pH-2II at a position corresponding to the last codon in clone pH-2I (Figure 8). Because of this insertion, the coding sequence of clone pH-2II is 13 codons longer than that of clone pH-2I (Figure 8). When these cDNA sequences are compared with that of gene 27.1, the 139 bp insertion is found in the same alignment as for clone pH-2II and corresponds to the last intervening sequence in clone 27.1 (Figure 9). This intervening sequence in 27.1 is also 139 bp in length and 90% homologous to the corresponding pH-2II cDNA sequence. We isolated a 138 bp Hph I–Pvu II fragment containing only sequences from the 139 bp insert of clone pH-2II except for 11 bp. This fragment does not contain the repetitive elements identified on the 3′ untranslated region of clone pH-2II (Steinmetz et al., 1981). When this fragment was used as a probe in Southern blot analysis of five different genomic clones, hybridization was noted for each of the five clones, suggesting that this sequence is contained in all of these genomic clones. Moreover, this intervening sequence has upstream and downstream RNA splicing signals at its 5′ and 3′ ends. Several explanations for the synthesis of pH-2I-like and pH-2II-like mRNAs are possible. It is conceivable that the pH-2II cDNA clone was derived from a partially spliced mRNA with the last intervening sequence still present. Clone pH-2II would then represent an intermediary splicing product and not a functional message. Alternatively, the retention of the last intervening sequence could also be of physiological significance. A differential RNA splicing mechanism might operate at the 3′ end of class I genes allowing the synthesis of proteins with markedly dif-

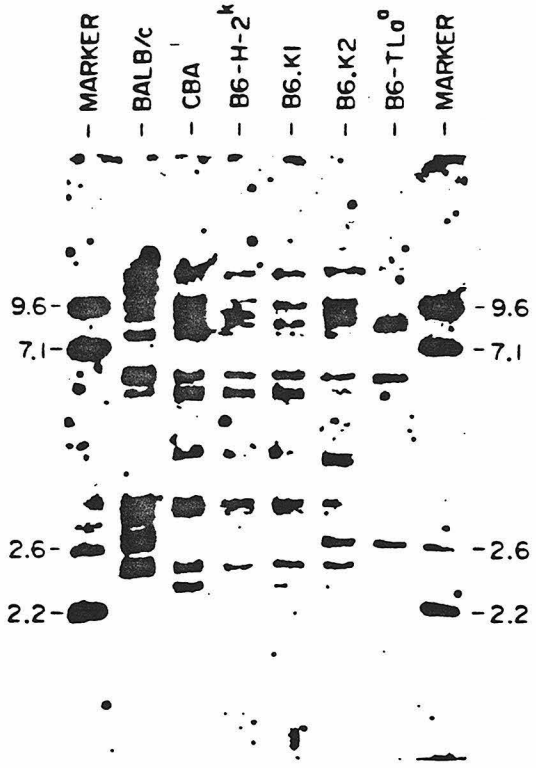


Figure 6. Analysis of the DNAs from Six Different Inbred Strains of Mice by Southern Blot Hybridization

Five micrograms each of liver DNA isolated from the six mouse inbred strains were completely digested with Bam HI, separated on a 0.5% agarose gel, transferred to a nitrocellulose filter and hybridized with the subclone pH-2IIa as a probe (Figure 2A). This clone was derived from pH-2II by subcloning the coding region and the portion of the 3′ untranslated region devoid of the repetitive sequences into pBR322 (Steinmetz et al., 1981). Hybridization conditions, washing and the hybridization marker were as described previously (Steinmetz et al., 1981). The sizes of the marker DNA fragments are given in kilobases.

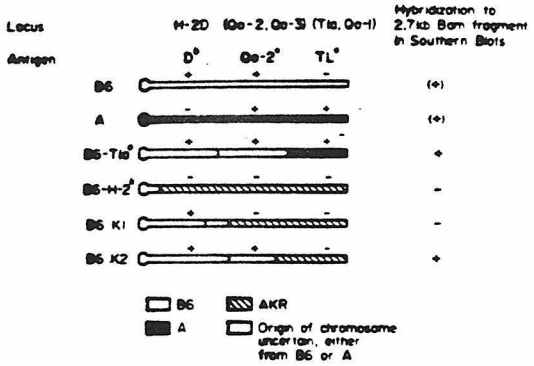


Figure 7. Mapping of Gene 27.1 by Southern Blot Analysis of Recombinant Congenic Strains of Mice

A limited map of mouse chromosome 17 is shown. The origin of the chromosomes is indicated by the various shading or stippling. The presence or absence of a given antigen (phenotype) is indicated by (+) or (−). B6-Tla^a is a congenic strain with the Tla region of A on a B6 background. B6.K1 and B6.K2 are recombinant strains derived from B6-Tla^a and B6-H-2^a. The presence or absence of the 2.7 kb Bam fragment in the Southern blot experiments (Figure 6) is indicated by (+) or (−). Brackets for B6 and A indicate that the Southern blot experiments (not shown in Figure 6) were performed with the substrain B10, and A/J, which is of the same haplotype as B6 and A/A, respectively. The results map the gene 27.1 to the right of the recombination event in B6.K1 and to the left of the recombination event in B6.K2. Figure adapted from Flaherty (1980).

ferent cytoplasmic domains (see Early et al., 1980). In this respect, it is interesting to note that the C-terminal portion of the K^b molecule is strikingly homologous to the translated sequence of gene 27.1 except for the last seven amino acids in this polypeptide chain (Figure 4). The breakpoint between the homologous C-terminal sequence and the last seven amino acids occurs precisely at the boundaries of the seventh and eighth exons, where the hypothetical splicing event would occur.

Genes for Transplantation Antigens and Antibody Molecules Are Homologous in Their General Structural Features

A computer comparison of the exons of gene 27.1 with exons of the immunoglobulin C_μ gene confirmed our previous finding for the cDNA clones (Steinmetz et al., 1981) of striking homology (up to 50%) between the DNA sequences of the third external domain of transplantation antigens and the constant region domains of immunoglobulin molecules. The availability of the 27.1 genomic sequence allowed us to extend the sequence comparison into the intervening sequences flanking the third external domain of gene 27.1. When compared with the fourth constant region domain of the μ gene with its flanking sequences, moderate homology of 35% was found for the first 75 bp of the 5′ flanking sequence of the third external domain without placing any sequence gaps into the two sequences compared (not shown). No significant homology was found for the 3′ flanking sequence. In contrast to the third external domain, the exons for the first and second external domains of the transplantation antigens do not show significant homology to immunoglobulins and show only marginal homology to one another (34% homology between the exons coding for the first and second external domains, 30% homology between those coding for the first and third external domains and 29% homology between those coding for the second and third external domains). The genes for transplantation antigens and antibodies both show a striking correlation between structural domains of the proteins and the exons for the corresponding genes.

The class I genes of the mouse constitute a multigene family with several interesting features. One such feature is that they appear to be distributed

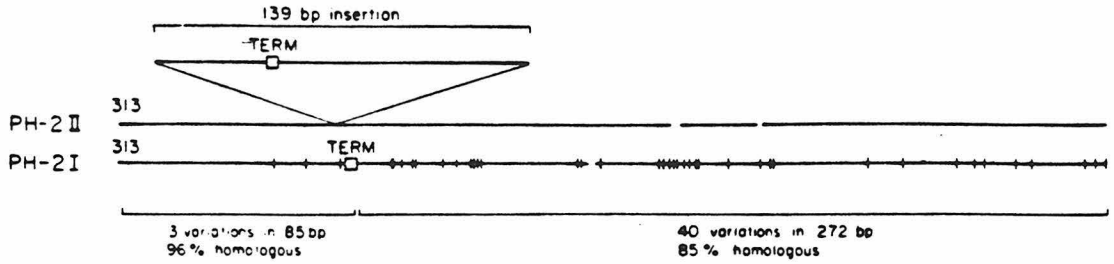


Figure 8 Schematic Comparison of the DNA Sequences for the 3' Regions of the cDNA Clones pH-2I and pH-2II

The sequences present in clone pH-2I and pH-2II are homologous except for a 139 bp insertion in clone pH-2II (Steinmetz et al., 1981 and Figure 5). The termination codons are boxed. Clone pH-2II has two sequence gaps, one of three nucleotides and the other of one nucleotide, and clone pH-2I has a single gap of two nucleotides that were introduced to achieve maximum homology. Nucleotides in pH-2I that differ from the pH-2II sequence are indicated by vertical bars.

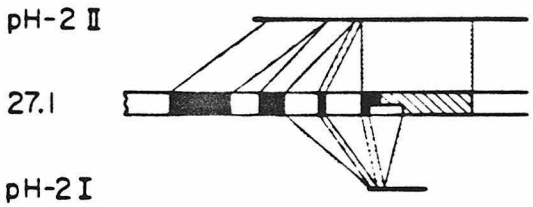


Figure 9 Two Hypothetical RNA Splicing Patterns at the 3' End of Gene 27.1 That May Generate pH-2II-Like and pH-2I-Like Messenger RNA Molecules

For gene 27.1, four exons encoding the third external domain, the transmembrane domain and the cytoplasmic domain (black boxes) as well as the first part of the 3' untranslated region (hatched box) are shown. The pH-2I-like mRNAs would be generated by an RNA splice that would remove the 139 bp intron indicated by the inserted open box, whereas in pH-2II-like mRNAs, this intron would not be removed.

across as much as 1.3 centiMorgans of DNA (several million nucleotides) (Boyse et al., 1965), if the Tla genes are homologous to those encoding transplantation antigens (Figure 1). If the Tla genes are homologous to the class I genes, perhaps the boundaries of the H-2 complex should be enlarged to include the Tla complex. Another feature is that the class I gene family is interrupted by the class II and class III gene families. It will be interesting to determine whether two or more of these three gene families share a common ancestor. Since the class I gene family is contiguous in other mammals, perhaps the organization of the H-2 complex in the mouse reflects a high level of evolutionary gene duplication, deletion and rearrangement in this gene family. Finally, the class I and antibody gene families appear to be members of a supergene family as reflected in their homology and their similar organizational features. These gene families use distinct regulatory strategies, since DNA rearrangements and somatic mutation occur in the antibody but probably not in the class I gene family. It will be interesting, however, to determine whether there are still other types of regulatory mechanisms shared by these two evolutionarily related gene families.

Experimental Procedures

Materials

Restriction enzymes were purchased from New England BioLabs, Bethesda Research Laboratories and Boehringer Mannheim. The 12 nucleotide primer used for DNA sequencing was obtained from Collaborative Research and the ddNTPs from Collaborative Research and P-L Biochemicals. E. coli DNA polymerase I large fragment was obtained from Boehringer Mannheim, T4 DNA polymerase from P-L Biochemicals, E. coli strain JM103 and phage M13mp7 from Bethesda Research Laboratories, T4 DNA ligase from New England BioLabs and α-^{32}P-dCTP (10 mCi/ml, >400 Ci/mmole) from Amersham. The E. coli strain JM101 was obtained from J. Messing, and the phage M13mp2 was obtained from J. Drouin.

DNA Sequence Analysis

A 3.6 kb Bgl II and a 2.7 kb Bam HI fragment including the coding regions were isolated from clone 27.1 by digestion of 50 μg phage DNA with the appropriate enzyme, gel electrophoresis on a 0.5% agarose gel and electroelution of the desired fragment followed by BD-cellulose purification. For cloning and sequencing by the M13 dideoxy sequencing technique (Sanger et al., 1980; Messing et al., 1981), we digested the isolated fragments with a number of restriction enzymes that are known to cut DNA frequently. The resulting fragment mixtures were cloned into M13 vectors and sequenced according to the procedures described by Sanger et al. (1980) and in the following protocol. Sequences were aligned by a computer search program written by T. Hunkapiller. During the course of this work, we found it easier and faster to cleave the large DNA fragments in only a few places, using infrequent cutting restriction enzymes, and to clone the resulting, relatively long DNA fragments to get extended stretches of sequence. If certain fragments were not found in the clone collections, they were individually isolated by gel electrophoresis and then cloned and sequenced. For the 2.7 kb Bam fragment, we used as frequent cutters the enzymes Alu I, Hae III, Hinf I, Rsa I and Sau 3A and as infrequent cutters the enzymes Pvu II and Sst I. The 3.6 kb Bgl II fragment was cleaved with Alu I, Hae III, Hinf I, Rsa I and Sau 96I as frequent cutters and for the generation of long fragments with mixtures of Bam HI + Pst I + Xba I and Bam HI + Pvu II.

As a cloning vector, we first used M13mp7 with the E. coli strain JM103 as host but switched later to M13mp2 with JM101 as host. About 50% of the mp7 clones that we picked gave faint sequencing patterns that were not readable. This was not observed with the mp2 vector. For cloning, the Eco RI–cleaved M13mp2 DNA and the fragments to be cloned were made flush-ended by incubation with T4 DNA polymerase and then were blunt-end ligated (Wartell and Reznikoff, 1980). Phage clones containing genomic DNA were identified by plaque hybridization (Benton and Davis, 1977) with clone 27.1 used as a probe.

For sequencing, one fortieth of the DNA isolated from a 1.5 ml culture (Sanger et al., 1980) was mixed with 2.5 ng of a 12 nucleotide

primer DNA fragment, 2.5 μCi of α-^{32}P-dCTP (Anderson et al., 1980) and 0.25 unit of E. coli DNA polymerase I large fragment in a final volume of 5 μl in 7 mM Tris–HCl (pH 7.5), 7 mM $MgCl_2$, 5 mM dithiothreitol, and 50 mM NaCl. Individual reactions were supplemented with the following concentrations of unlabeled nucleotides: A reaction, 1.6 μM dATP and 100 μM ddATP; C reaction, 50 μM ddCTP; G reaction, 1.6 μM dGTP and 100 μM ddGTP; T reaction, 1.6 μM dTTP and 400 μM ddTTP. Each reaction also contained each of the remaining three dNTPs at the concentration of 25–33 μM. The reactions were incubated for 15 min at 30°C, then chased by the addition of 1 μl of a solution containing all four dNTPs, each at a concentration of 300 μM. After another 15 min incubation at 30°C, 10 μl of a solution containing 95% formamide, 10 mM EDTA (pH 8.0), 10 mM NaOH, 0.03% xylene cyanol and 0.03% bromophenol blue were added and the samples were stored frozen. For electrophoresis, the samples were treated for 3 min at 90°C and quick-cooled in ice-water and 2–4 μl were loaded on 40 cm 8% and 80 cm 4% or 5% acrylamide sequencing gels (Smith and Calvo, 1980). Exposure was usually overnight with intensifying screen. To read individual sequences beyond position 200, it was usually necessary to decrease the ddNTP concentration in the reactions by a factor of two to eight.

As indicated in Figure 2C, four regions of the 2.7 kb Bam fragment and a part of the 3.8 kb Bgl II fragment also were sequenced by the chemical degradation method (Maxam and Gilbert, 1980). Fragments were labeled at their 3′ ends with E. coli DNA polymerase I large fragment (Smith and Calvo, 1980) and then divided by secondary restriction enzyme cleavage. For the G + A reaction, the procedure described by Gray et al. (1978) was used.

Isolation of Eucaryotic DNA and DNA Blot Hybridization

DNA was isolated from frozen mouse livers according to the procedure described by Blin and Stafford (1976), except that the tissues were pulverized in a mortar in the presence of liquid nitrogen, and two chloroform–isoamylalcohol (24:1) extractions were added after the phenol extractions. DNA blots (Southern, 1975) were prepared and hybridized according to the conditions previously described (Steinmetz et al., 1981).

Acknowledgments

We thank Tim Hunkapiller for computer programs, Michael Douglas for computer graphics, Bernita Larsh for preparation of the manuscript and Karyl Minard for excellent technical assistance. We also thank Dr. J. Drouin for his advice regarding dideoxy sequencing in the M13 system. M. S. is the recipient of a fellowship from the Deutsche Forschungsgemeinschaft and K. W. M. and J. G. F. are supported by National Institutes of Health postdoctoral fellowships. B. T. S. is a National Science Foundation fellow. E. A. B. is American Cancer Society Research Professor of Cell Surface Immunogenetics. 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 July 6, 1981

References

Anderson, S., Gait, M. J., Mayol, L. and Young, I. G. (1980). A short primer for sequencing DNA cloned in the single-stranded phage vector M13mp2. Nucl. Acids Res. *8*, 1731–1743.

Benton, W. D. and Davis, R. W. (1977). Screening λgt recombinant clones by hybridization to single plaques in situ. Science *196*, 180–182.

Blin, N. and Stafford, D. (1976). A general method for isolation of high molecular weight DNA from eukaryotes. Nucl. Acids Res. *3*, 2303–2308.

Boyse, E. A., Old, L. J. and Luell, S. (1964). Genetic determination of the TL (thymus leukemia) antigen in the mouse. Nature *201*, 779.

Boyse, E. A., Old, L. J. and Stockert, E. (1965). The TL (thymus leukemia) antigen. A review. In Immunopathology, P. Grabar and P. A. Miescher, eds. (New York: Grune and Stratton), pp. 23–40.

Brégégère, F., Abastado, J. P., Kvist, S., Rask, L., Lalanne, J. L., Garoff, H., Cami, B., Wiman, K., Larhammar, D., Peterson, P. A., Gachelin, G., Kourilsky, P. and Dobberstein, B. (1981). Structure of C-terminal half of two H-2 antigens from cloned mRNA. Nature *292*, 78–81.

Cami, B., Brégégère, F., Abastado, J. P. and Kourilsky, P. (1981). Multiple sequences related to classical histocompatibility antigens in the mouse genome. Nature *291*, 673–675.

Coligan, J. E., Kindt, T. J., Uehara, H., Martinko, J. and Nathenson, S. G. (1981). Primary structure of a murine transplantation antigen. Nature *291*, 35–39.

Davis, M. M., Calame, K., Early, P. W., Livant, D. L., Joho, R., Weissman, I. L. and Hood, L. (1980). An immunoglobulin heavy-chain gene is formed by at least two recombinational events. Nature *283*, 733–739.

Dobberstein, B., Garoff, H., Warren, G. and Robinson, P. J. (1979). Cell-free synthesis and membrane insertion of mouse H-2D^d histocompatibility antigen and β2-microglobulin. Cell *17*, 759–769.

Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R. and Hood, L. (1980). Two mRNAs can be produced from a single immunoglobulin μ gene by alternative RNA processing pathways. Cell *20*, 313–319.

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

Flaherty, L. (1980). The Tla region antigens. In The Role of the Major Histocompatibility Complex in Immunology, M. E. Dorf, ed. (New York: Garland STPM Press), pp. 33–57.

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

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

Jordan, B. R., Brégégère, F. and Kourilsky, P. (1981). Human HLA gene segment isolated by hybridization with mouse H-2 cDNA probes. Nature *290*, 521–523.

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

Kvist, S., Brégégère, F., Rask, L., Cami, B., Garoff, H., Daniel, F., Wiman, K., Larhammar, D., Abastado, J. P., Gachelin, G., Peterson, P. A., Dobberstein, B. and Kourilsky, P. (1981). cDNA clone coding for part of a mouse H-2^d major histocompatibility antigen. Proc. Nat. Acad. Sci. USA *78*, 2772–2776.

Lewin, B. (1980). Alternatives for splicing: recognizing the ends of introns. Cell *22*, 324–326.

Lopez de Castro, J. A., Orr, H. T., Robb, R. J., Kostyk, T., Mann, D. L. and Strominger, J. L. (1979). Complete amino acid sequence of a papain-solubilized human histocompatibility antigen, HLA-B7. I. Isolation and amino acid composition of fragments and of tryptic and chymotryptic peptides. Biochemistry *18*, 5704–5711.

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

Messing, J., Crea, R. and Seeburg, P. H. (1981). A system for shotgun DNA sequencing. Nucl. Acids Res. *9*, 309–321.

Michaelson, J., Flaherty, L., Vitetta, E. and Poulik, M. (1977). Molecular similarities between the Qa-2 alloantigen and other gene products of the 17th chromosome of the mouse. J. Exp. Med. *145*, 1066–1070.

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 *282*, 266–270.

Ploegh, H. L., Cannon, L. E. and Strominger, J. L. (1979). Cell-free

translation of the mRNAs for the heavy and light chains of HLA-A and HLA-B antigens. Proc. Nat. Acad. Sci. USA 76, 2273–2277.

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.

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

Proudfoot, N. J. and Brownlee, G. G. (1976). 3' Non-coding region sequences in eukaryotic messenger RNA. Nature *263*, 211–214.

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

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.

Sotoski, M. J., Uhr, J. W., Flaherty, L. and Vitetta, E. S. (1981). Qa-2, H-2K, and H-2D alloantigens evolved from a common ancestral gene. J. Exp. Med. *153*, 1080–1093.

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 78, 616–620.

Southern, E. M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. *98*, 503–517.

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

Stanton, T. H. and Hood, L. (1980) Biochemical identification of the Qa-1 alloantigen. Immunogenetics *11*, 309–314.

Steinmetz, M., Frelinger, J. G., Fisher, D., Hunkapiller, T., Pereira, D., Weissman, S. M., Uehara, H., Nathenson, S. and Hood, L. (1981). Three cDNA clones encoding mouse transplantation antigens: homology to immunoglobulin genes. Cell *24*, 125–134.

Strominger, J. L., Orr, H. T., Parham, P., Ploegh, H. L., Mann, D. L., Biofsky, 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.

Wartell, R. M. and Reznikoff, W. S. (1980) Cloning DNA restriction endonuclease fragments with protruding single-stranded ends. Gene *9*, 307–319.

Appendix B

DNA Sequence of the H-2L^{d} Gene.

The H-2L^{d} gene was the first transplantation antigen gene to be sequenced. The publication contained in this appendix originally appeared in Science.

Reprint Series
5 February 1982, Volume 215, pp. 679–682

SCIENCE

DNA Sequence of a Gene Encoding a BALB/c Mouse L^d Transplantation Antigen

Kevin W. Moore, Beverly Tayor Sher, Y. Henry Sun, Kurt Andrew Eakle, and Leroy Hood

DNA Sequence of a Gene Encoding a BALB/c Mouse L^d Transplantation Antigen

Abstract. *The sequence of a gene, denoted 27.5, encoding a transplantation antigen for the BALB/c mouse has been determined. Gene transfer studies and comparison of the translated sequence with the partial amino acid sequence of the L^d transplantation antigen establish that gene 27.5 encodes an L^d polypeptide. A comparison of the gene 27.5 sequence with several complementary DNA sequences suggests that the BALB/c mouse may contain a number of closely related L-like genes. Gene 27.5 has eight exons that correlate with the structural domains of the transplantation antigen.*

The transplantation or class I antigens of the major histocompatibility complex (MHC) are present on the surface of all mammalian somatic cells and play a key role in T cell immunosurveillance (*1*, *2*). In the inbred BALB/c mouse, three genes for transplantation antigens, D, L, and R, are closely linked and these are separated from a fourth class I gene, K, by approximately 0.3 centimorgan (*3*). We cloned 30 to 40 genes of the BALB/c mouse ($H\text{-}2^d$ haplotype) that are homologous to complementary DNA (cDNA) probes for transplantation antigens (*4*) and have demonstrated by gene transfer experiments that one of these clones (27.5) encodes an L^d transplantation antigen (*5*). We have determined the nucleotide sequence of the L^d gene present in clone 27.5.

The genomic clone 27.5 was isolated as previously described (*4*) from an amplified library of BALB/c sperm DNA cloned in the λ vector Charon 4A. The nucleotide sequence of gene 27.5 was determined by the dideoxy sequencing technique with M13mp2 as cloning vector (*4*, *6*, *7*). The sequencing strategy is shown in Fig. 1. The exons of gene 27.5 were defined on the basis of homology to available amino acid sequences for transplantation antigens (*8*, *9*), to several DNA sequences of cDNA's encoding transplantation antigens (*10*, *11*), and to a genomic class I clone 27.1, which bears a pseudogene mapping to the Qa-2,3 re-

0036-8075/82/0205-0679$01.00/0

gion (*4*). The DNA sequence of gene 27.5 is given in Fig. 2.

Gene 27.5 has eight exons whose boundaries correspond precisely to those determined earlier for the class I pseudogene 27.1 (Fig. 1c) (*4*). Apart from introns 1 and 3, the lengths of the introns in genes 27.5 and 27.1 are similar. Exon 1 encodes the signal peptide; exons 2 to 4, the three external domains; exon 5, the transmembrane segment; and exons 6 to 8, the cytoplasmic domain. There is a striking correlation between the discrete exon boundaries and the structural domains of the transplantation antigen (*4*). All of the exon-intron boundaries have the consensus upstream or downstream RNA splicing signals (*12*). Gene 27.5 appears to be a functional gene by sequence analysis in that it lacks any obvious elements that would render it a pseudogene (for example, termination codons or inappropriate reading frame shifts). This conclusion is supported by our gene transfer studies (*5*).

The translated sequence of gene 27.5 is identical to the amino acid sequence of the L^d molecule at 77 of 77 positions that can be compared (Table 1) (*13*). These comparisons include residues in the first, second, and third external domains. Although the paucity of amino acid sequence data permits us to compare only 21 percent (77 of 358 positions) of the gene and protein sequences, these comparisons support the conclusion reached in gene transfer studies that clone 27.5 contains an L^d gene (*5*).

There is one striking observation that can be made in comparing the sequences of gene 27.5 and of two cDNA clones, pH-2II (*10*) and pH-2^d-3 (*11*), whose translated sequences are identical with the available partial L^d amino acid sequence. Gene 27.5 is identical to cDNA clone pH-2^d-3 at 467 of 470 positions compared. Two of the sequence differences are in the first 60 nucleotides of the fourth exon and lead to one codon substitution. The other difference, which also leads to a codon substitution, is in the fifth exon. Likewise, the exons of gene 27.5 are identical to clone pH-2II at 514 of 520 nucleotide positions. The differences are clustered in the first 45 nucleotides of the fourth exon and lead to five codon substitutions. There are also six nucleotide differences between gene 27.5 and clone pH-2II of the 3′ untranslated region of 587 nucleotides compared. It is interesting to note that the differences in coding sequence are clustered at the beginning of the fourth exon, because this exon encodes the third external domain of the L^d polypeptide, and this domain is the most conserved region of the genes encoding transplantation antigens (*14*).

The genomic clone 27.5 was derived from BALB/c (H-2^d haplotype) sperm DNA. The cDNA clone pH-2II was derived from a BALB/c tumor cell line (*10*), and the cDNA clone pH-2^d-3 was derived from a DBA/2 (H-2^d haplotype) lymphoma line (*11*). It is unlikely that all of the differences in these class I sequences arose from cloning or sequencing artifacts. Accordingly, mice of the H-2^d haplotype may have three closely related class I genes that encode L^d-like polypeptides. The three L-like genes might arise from genetic polymorphism at a single locus within mice of the H-2^d haplotype or from duplicated genes.

The possibility that these different clones represent duplicated genes is attractive in view of historical precedent and of recent evidence. By serological analysis, the D end of the H-2 complex initially appeared to have a single D gene (*2*). Subsequently, more refined serological analyses have suggested that this region encodes three closely linked genes for transplantation antigens—D, L, and R (*3*). Cosmid clones that have been isolated from BALB/c DNA contain three L-like genes according to restriction map analyses (*14*). Gene transfer studies have demonstrated that at least one of these cosmid genes encodes an L^d polypeptide (*15*). It will be interesting to determine whether the remaining two of these putative L-like genes also are expressed as class I polypeptides reacting with the monoclonal antibodies to the L^d antigen.

A comparison of the translated sequences of pseudogene 27.1 and the L^d gene 27.5 with the amino acid sequence of the mouse K^b molecule (Table 1) shows that pseudogene 27.1 appears to have diverged significantly more from the translated L^d gene than the K^b molecule has (Table 1). The difference between 27.1 and the other class I genes might arise because the genes encoding Qa antigens diverged from those encoding the classical transplantation antigens before the divergence of the individual class I genes. Alternatively, this divergence may reflect changes accumulating in the pseudogene, which has presumably been released from selective pressure. In addition, the exons of pseudo-

Table 1. Homology of the exons of the translated L^d gene (27.5) to the corresponding regions of the translated pseudogene 27.1 (*4*) and to the K^b (*12*) and L^d (*13*) molecules. The homology of the translated portion of each exon to the corresponding portion of the K^b and L^d sequences, as well as to the translated 27.1 pseudogene exons (*4*), is shown.

Exon	Homology (%)		
	K^b	L^{d*}	27.1
Leader (exon 1)			76
First domain (exon 2)	84	100	71
Second domain (exon 3)	80	100	76
Third domain (exon 4)	88	100	89
Transmembrane domain (exon 5)	73		69
Cytoplasmic exon			
First (exon 6)	100		73
Second (exon 7)	64		69
Third (exon 8)	0		100

*Thirty-four of 90 positions were compared in exon 2, 28 of 92 in exon 3, and 15 of 92 in exon 4.

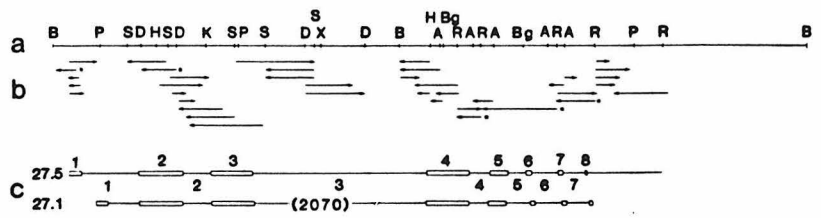


Fig. 1. The organization and sequencing strategy for gene 27.5. (a) A restriction map of clone 27.5. This map was generated by Southern blot analyses of clone 27.5 cleaved with various restriction enzymes and probed with selected M13 clones into which fragments of gene 27.5 had been inserted (see below). The restriction sites are designated as follows: *B*, Bam HI; *P*, Pst I; *S*, Sau 3A; *D*, Dde I; *K*, Kpn I; *Bg*, Bgl II; *R*, Rsa I; *H*, Hinf I; *X*, Xba I; and *A*, Alu I. (b) Sequence strategy for gene 27.5. Each arrow represents the sequence of an M13 clone. The M13 clones indicated by asterisks were used as probes in Southern blot analyses to generate the restriction map. At positions where clones do not overlap, alignment was determined by restriction mapping or by homology with cDNA clone pH-2II. (c) Organization of genes 27.5 and 27.1. Exons are represented by boxes and introns by lines. Both exons (above boxes) and introns (below lines) are numbered.

gene 27.1 are homologous to those of the L^d gene 27.5. This homology suggests a common evolutionary origin. If gene 27.1 is a Qa-2,3 pseudogene [see (*4*) for discussion], then this homology suggests that the Qa antigens are class I molecules and that the Tla complex should be considered a part of the H-2 complex. This conclusion is supported by observations that the TL and Qa antigens resemble the classical transplantation antigens in size, peptide map profiles, and their association with β_2-microglobulin (*16–19*).

We have used a computer-generated dot matrix to analyze the homology relationships of the DNA sequences from genes 27.1 and 27.5 (Fig. 3). This analysis compares every hexamer of gene 27.1 (vertical axis) against every hexamer of gene 27.5 (horizontal axis) and places a dot in the two-dimensional matrix at positions where at least five of six nucleotides are identical. Homologies are displayed as diagonal lines in the matrix; nucleotide divergences are represented as gaps in the diagonal lines; and sequence insertions or deletions offset the diagonal lines (*20*). Four important points emerge from this analysis. (i) Genes 27.1 and 27.5 are quite homologous to one another, as can be seen by the strong diagonal lines. (ii) Five of seven introns (2, 4, 5, 6, and 7) are almost as highly conserved in both genes as the exons are; the exons exhibit 87 percent homology, and these introns ex-

```
                                                                                                  EXON 1
                                                                                       Met Ala Pro Arg Thr Leu Leu Leu Leu Leu
CCAGGGCGGATTGAGAGGGAAGACCACACACCCTGTGAGCTCACTGTGTCCAGTGAGTAGCTGCACTGGGGTCCACAGCTCACTCCAGGGATCACTCCCAGATGGGGGCG ATG GCT CCG CGC ACG CTG CTC CTG CTG CTG    140

Ala Ala Ala Trp Pro Asp Ser Asp Pro Arg G
GCG GCC GCC TGG CCC GAC TCA GAC CCG CGG G GTGAGTGSGGGTGCGGAGGGAAACGGCCTCTGCGGGGAGAGGGGCGGCAGGGGGAAGCGC ...GATCCCCTCGCCTCGCAGCCGCCCGGGGTTTGGTAGGAGGTCGG    278

                                                           EXON 2
                        ly Pro His Ser Met Arg Tyr Phe Glu Thr Val Ser Arg Arg Gly Leu Gly Glu Pro Arg Tyr Ile Ser Val Gly Tyr Val Asp Asn Lys Glu Phe
GGTCTCACCGCGCGCCCGCCCCAG GC CCA CAC TCG ATG CGG TAT TTC GAG ACC GTG TCC CGG CGC GGC CTC GGG GAG CCC CGG TAC ATC TCT GTC GGC TAT GTG GAC AAC AAG GAG TTC    387

Val Arg Phe Asp Ser Asp Ala Glu Asn Pro Arg Tyr Glu Pro Arg Ala Pro Trp Met Glu Gln Glu Gly Pro Glu Tyr Trp Glu Arg Ile Thr Gln Ile Ala Lys Gly Gln Glu
GTG CGC TTC GAC AGC GAC GCG GAG AAT CCG AGA TAT GAG CCG AGG GCG CCG TGG ATG GAG CAG GAG GGG CCG GAG TAT TGG GAG CGG ATC ACG CAG ATC GCC AAG GGC CAG GAG    511

Gln Trp Phe Arg Val Asn Leu Arg Thr Leu Leu Gly Tyr Tyr Asn Gln Ser Ala Gly G
CAG TGG TTC CGA GTG AAC CTG AGG ACC CTG CTC GGC TAC TAC AAC CAG AGC GCG GGC G GTGAGTGACCCCGCGTCGGAGGTCACGGCCCCTCCCTTTCCCGACACAGGGACGCTGACTTCGTACCCAAGT    641

                                                                                                   ly Thr His Thr Leu Gln Trp Met Tyr
CCGAGGTTCGGGAACAGAACGGACCCGGAACCAGTTTCCCTTTCAGTTTGGAGGAGTCGCGGCCGGGCGGCGGGGCCGGGGCGGGTGAGCGGGGCTGAACCGGGGTACCGCAG GC ACT CAC ACA CTC CAG TGG ATG TAC    780

                                                           EXON 3
Gly Cys Asp Val Gly Ser Asp Gly Arg Leu Leu Arg Gly Tyr Glu Gln Phe Ala Tyr Asp Gly Cys Asp Tyr Ile Ala Leu Asn Glu Asp Leu Lys Thr Trp Thr Phe Ala Asp
GGC TGT GAC GTG GGG TCG GAC GGG CGC CTC CTC CGC GGG TAC GAG CAG TTC GCC TAC GAC GGC TGC GAT TAC ATC GCC CTG AAC GAA GAC CTG AAA ACG TGG ACG TTC GCG GAC    894

Met Ser Ser Gln Ile Thr Arg Arg Lys Trp Glu Gln Ala Gly Ala Ala Glu Tyr Tyr Arg Ala Tyr Leu Glu Gly Glu Cys Val Glu Trp Leu His Arg Tyr Leu Lys Asn Gly
ATG TCG TCG CAG ATC ACC CGA CGC AAG TGG GAG CAG GCT GGT GCT GCA GAG TAT TAC AGG GCC TAC CTG GAG GGC GAG TGC GTG GAG TGG CTC CAC AGA TAC CTG AAG AAC GGG    1008

Asn Ala Thr Leu Leu Arg Thr A
AAT GCT ACG CTG CTG CGC ACA G GTGCAGGGGCCGCGGCGGGCAGGCTCCTCCCTCTCGCCTCGGGCTGGGGGCTCAGTCCTGGGGAAGAAGAAACCCTCAGCTGGGGTGATGCCCTGTCTCAGAGGGGAGAGAGTGACCCTG    1150

GGTTCCTGATGCCTCAGCACAGTGACTGCACTGACTCTCCCAGGGCTCAGCTTCTCCCTGGACAGTCCCAGCCTGTCTCAGGAGGGAAGGAGAGAATTTCCCTGAGGTAACAACAGCTGCTCCCTTCAGTTCCCCTGTAGCCTCTGTCAG    1300

CCATGGCCTCTCCCAGGCGGGTTCTCTGCCCACGCCACTGTCGGTGACACTGACTCCTGTCCTGGCTGATGTGTGTCAGCGCCTTACACCTCAGGACCGGAAGTCGCCTTACCTGATTGGAAACATGGACTCCTATACACTAGCCGTGTT    1450

GCCCCAGCTTCTAGAACTTTCCAGAGAATACATTCTCCCAGATCCCTCCCTGTCTGTGGGGTTTGCACCCCTTCGCACAACCTAATTCTCTCTATTCCTACAGTGGTGAATGGTCACATGAGCTCTTATGGGGTACSTGGAGGAATATAA    1600

ATCGAGAATTTTCTTTTTTGTTTTCTCTCTCTCTTTCTCTCTCTCTCTCTCTCTCTCTTTTCCGATTTTTTTGAGACCGGGTTTCTCTCTATAGCCCTSTGTCCTGACTCATTTTAGCCCTTTTCTAATTCACATCTTCTCT ...TT    1740

                                                                                                              sp Ser Pro Lys Ala
TGGTCACTAGTGCAATGACAGTGTAGTGTCAAATAGACACATAGTTCACTCTCATCATTGATTTAACTCAGTCTTGTGTAGATTTCAGTTTGTCTTGTTAATTGTGGAATTTCTTAAATCTTCCACACAG AT TCC CCA AAG GCA    1883

                                                           EXON 4
His Val Thr His His Pro Arg Ser Lys Gly Glu Val Thr Leu Arg Cys Trp Ala Leu Gly Phe Tyr Pro Ala Asp Ile Thr Leu Thr Trp Gln Leu Asn Gly Glu Glu Leu Thr
CAT GTG ACC CAT CAC CCC AGA TCT AAA GGT GAA GTC ACC CTG AGG TGC TGG GCC CTG GGC TTC TAC CCT GCT GAC ATC ACC CTG ACC TGG CAG TTG AAT GGG GAG GAG CTG ACC    2007

Gln Asp Met Glu Leu Val Glu Thr Arg Pro Ala Gly Asp Gly Thr Phe Gln Lys Trp Ala Ser Val Val Val Pro Leu Gly Lys Glu Gln Asn Tyr Thr Cys Arg Val Tyr His
CAG GAC ATG GAG CTT GTG GAG ACC AGG CCT GCA GGG GAT GGA ACC TTC CAG AAG TGG GCA TCT GTG GTG GTG CCT CTT GGG AAG GAG CAG AAT TAC ACA TGC CGT GTG TAC CAT    2121

Glu Gly Leu Pro Glu Pro Leu Thr Leu Arg Trp G
GAG GGG CTG CCT GAG CCC CTC ACC CTG AGA TGG G GTAAGGAGGGTGTGGGTGCAGAGCTGGGGTCAGGGAAAGCTGGAGCCTTCTGCAGACCCTGAGCTGGTCAGGGATGAGAGACTGGGGTCATAACCCTCACCTTC    2250

                                                           EXON 5
                        lu Pro Pro Pro Ser Thr Asp Ser Tyr Met Val Ile Val Ala Val Leu Gly Val Leu Gly Ala Met Ala Ile Ile Gly Ala Val Val Ala Phe Val
ATTTCCTGTACCTGTCCTTCCCAG AG CCT CCT CCG TCC ACT GAC TCT TAC ATG GTG ATC GTT GCT GTT CTG GGT GTC CTT GGA GCT ATG GCC ATC ATT GGA GCT GTG GTG GCT TTT GTG    2378

Met Lys Arg Arg Arg Asn Thr G
ATG AAG AGA AGG AGA AAC ACA G GTAAGAAAGGGNAGGGTCTGAGTTTTCTCTCAGCCTCCTTTAGAAGTGTCTCTGCTCATTAATGGGGAACACAGCCACACCCCACATTGCTACTGTCTCTAACTGGGTCTGCTGTCAGTT    2520

                                                           EXON 6
                                                  ly Gly Lys Gly Gly Asp Tyr Ala Leu Ala Pro G
CTGGGAARTTCCAGTGTCAAGATCTTCCTTGAACTCTCACAGCTTTXCTTTTCACAG GT GGA AAA GGA GGG GAC TAT GCT CTG GCT CCA G GTTANTGTGGGGACAGGATNGTTCTGGGGGACATTGGAGTGAAGTTG    2657

                                                                                                                  EXON 7
                                                                                                      ly Ser Gln Ser Ser Glu
GAGATGATGGGAGCTCTGGGAATCCATAATAGCTCCTCCAGAGAAATCTTCTAGGGGCCTGAGTTGTGCCATGAAGTGAATACATTCATGTACATATGCATATACATTTGTTTTGTTTTACCCTAG GC TCC CAG AGC TCT GAA    2800

Met Ser Leu Arg Asp Cys Lys A
ATG TCT CTC CGA GAT TGT AAA G GTGACACTCTAGGGTCTGATTGGGGAGGGGCAATGTGGACATGATTGGGTTTCAGGGACTCCCAGAATCTCCTGAGAGTGAGTGGTGGGTTGCTGGAATGTTGTCTTCACAGTGATGGTT    2942

                    EXON 8
                    la Trm
CATGACTCTCATTCTCTAG CG TGA AGACAGCTGCCTGGACTGTACTGAGTGACAGACGATGTGTTCAGGTCTCTCCTGTGACATCCAGAGCCCTCAGTTCTCTTTACACAZCATTGTCTGATGTTCCCTGTGAGCTTGGGTTCAGTG    3080

TGAAGAACTGTGGAGCCCAGCCTGCCCTGCACACCAGGACCCTATCCCTGCACTGCCCTGTGTTCCCTTCCATAGCCAACCTTGCTGCTCCAGTCAAACACTGGGGGACATCTGCATCCTGTAAGCTCCATGCTACCCTGAGCTGCAGCT    3230

CCTCACTTCCACACTGAGAATAATAATTTGAATGTGGGTGGCTGGAGAGATGGCTCAGCGCTGACTGCTCTTCCAAAGGTCCTGAGTTCAAATCCCAGCAACACATGGTGGCTCACAACCATCTGTAATGGTATCTAACACCCTCTTCTG    3380

CAGTGTCTGAAGACAGCTACAGTGT    3414
```

Fig. 2. Nucleotide sequence of gene 27.5. The amino acid translation of each exon is given above the nucleotide sequence. Certain amino acid residues are encoded by two exons and are so indicated by splits in the three-letter amino acid code. Triple dots indicate gaps in intron sequences. Gaps in introns 1 and 3 are each approximately 200 nucleotides in length. The ambiguous base code is as follows: *P*, either A or G; *R*, either A or T; *Z*, A, T, or C; *Y*, either C or T; *S*, either G or C; *X*, A, T, or G.

hibit 80 percent homology. (iii) Two of the introns, 1 and 3, exhibit extensive regions of nonhomology in these two genes. (iv) The homology in the 3′ end of these genes terminates almost precisely at the end of the 3′ untranslated region as defined by comparison with the L^d-like cDNA clone pH-2II (*10*). Thus the exons and many of the introns of these class I genes are conserved. However, extensive divergences both in size (Fig. 1c) and sequence are seen in two introns and beyond the 3′ untranslated region of these genes. To what extent these divergences reflect a lack of selective pressure operating in these regions or the rapid changes that may occur in a pseudogene cannot be determined at this time.

Two highly repetitive sequence elements homologous to the human Alu sequence (*21*) have been found in the third intron of pseudogene 27.1 (*4*). Since Alu-like repetitive sequences appear to be ubiquitous in mammals and are transcribed in many cells, it has been postulated that they may play some undefined role in gene expression (*22*). The Alu-like sequences are not present in the third intron of gene 27.5. Hence the presence of these Alu-like sequences in intron 3 does not play a role in regulating the expression of gene 27.5. Furthermore, the Alu-like sequences occur in the third intron of gene 27.1 at precisely the boundaries of a large region of nonhomology (1000 nucleotides) in the comparison with the third intron in gene 27.5 (Fig. 3). Thus it is tempting to speculate that the Alu-like sequences, which have the characteristics of transposons (*22*), may be responsible for the insertion of a large region of foreign DNA in intron 3 of gene 27.1. This is consistent with the observation that the third intron of 27.1 is approximately 1000 nucleotides larger than the third intron of 27.5. It will be interesting to determine whether Alu-like sequences are present in other class I genes.

In summary, we have determined the coding sequence for an L^d gene of the mouse major histocompatibility complex. Thus, an H-2 gene corresponding to a serologically defined protein product has been identified and characterized. Cloning and DNA sequence studies suggest that there may be three (or more) L^d-like genes. It will be interesting to determine whether these putative L-like genes are codominantly expressed on all cells, or whether these genes are expressed in a tissue-specific manner such as is seen with the Qa and TL antigens. It appears that, in the future, other D-end functional genes will be defined by the strategies of gene cloning, gene transfer, and sequence analysis.

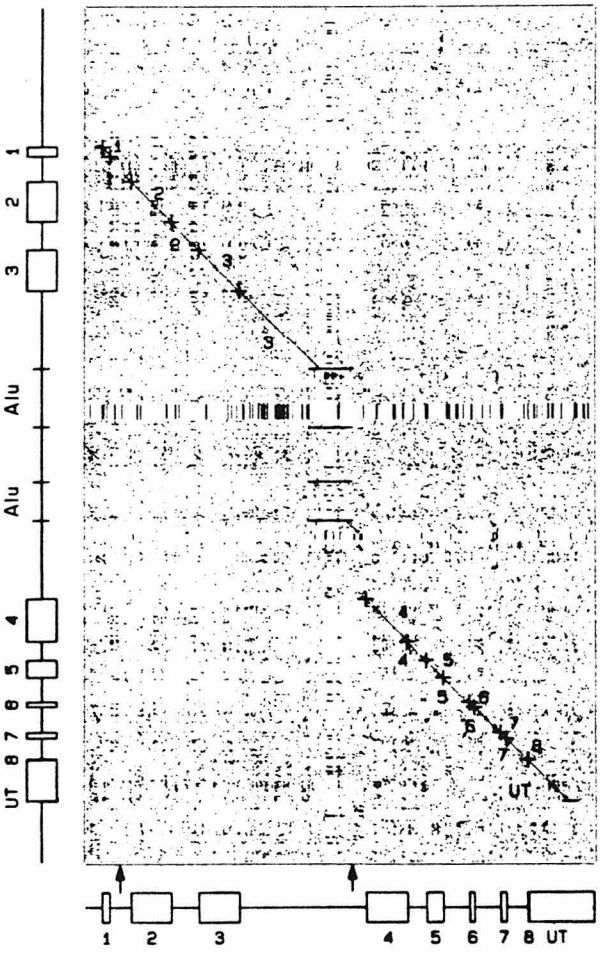


Fig. 3. Dot matrix homology analysis of gene 27.5 and pseudogene 27.1. The exon and intron structures of the two genes are shown along the sides of the figure. The positions of intron-exon boundaries are indicated by horizontal lines for 27.1 and vertical lines for 27.5 along the line of homology. Introns are numbered below the line of homology and exons are numbered above the line of homology. The positions of the sequence gaps in gene 27.5 are indicated by arrows. The positions of the Alu-like repetitive elements in gene 27.1 are indicated by horizontal lines along the line of homology and in intron 3 of pseudogene 27.1.

KEVIN W. MOORE
DNAX Research Institute, Palo Alto, California 94304
BEVERLY TAYLOR SHER
Y. HENRY SUN
KURT ANDREW EAKLE
LEROY HOOD*
Division of Biology, California Institute of Technology, Pasadena 91125

References and Notes

1. J. Klein, *Science* **203**, 516 (1979).
2. G. D. Snell, J. Dausset, S. Nathenson, *Histocompatibility* (Academic Press, New York, 1976).
3. T. H. Hansen, K. Ozato, M. R. Melino, J. E. Coligan, T. J. Kindt, J. J. Jandinski, D. H. Sachs, *J. Immunol.* **126**, 1713 (1981).
4. M. Steinmetz, K. W. Moore, J. G. Frelinger, B. T. Sher, F.-W. Shen, E. A. Boyse, L. Hood, *Cell* **25**, 683 (1981).
5. R. S. Goodenow, M. McMillan, A. Örn, M. Nicolson, N. Davidson, J. A. Frelinger, L. Hood, *Science* **215**, 677 (1982).
6. F. Sanger, A. R. Coulson, B. G. Barrell, A. J. H. Smith, B. A. Roe, *J. Mol. Biol.* **143**, 161 (1980).
7. J. Messing, R. Crea, P. H. Seeburg, *Nucleic Acids Res.* **9**, 309 (1981).
8. J. A. Lopez de Castro, H. T. Orr, R. J. Robb, T. Kostyk, D. L. Mann, J. L. Strominger, *Biochemistry* **18**, 5704 (1979).
9. J. E. Coligan, T. J. Kindt, H. Uehara, J. Martinko, S. G. Nathenson, *Nature (London)* **291**, 35 (1981).
10. M. Steinmetz *et al.*, *Cell* **24**, 125 (1981).
11. F. Bregegere *et al.*, *Nature (London)* **292**, 78 (1981).
12. B. Lewin, *Cell* **22**, 324 (1980).
13. J. E. Coligan, T. J. Kindt, R. Nairn, S. G. Nathenson, D. H. Sachs, T. H. Hansen, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1134 (1980).
14. M. Steinmetz, A. Winoto, K. Minard, L. Hood, *Cell*, in press.
15. R. Goodenow, unpublished results.
16. J. Michaelson, L. Flaherty, E. Vitetta, M. D. Poulik, *J. Exp. Med.* **145**, 1066 (1977).
17. L. Flaherty, *The Role of the Major Histocompatibility Complex in Immunology* (Garland, New York, 1980).
18. T. H. Stanton and L. Hood, *Immunogenetics* **11**, 309 (1980).
19. M. J. Soloski, J. W. Uhr, L. Flaherty, E. S. Vitetta, *J. Exp. Med.* **153**, 1080 (1981).
20. T. Hunkapiller and L. Hood, in preparation.
21. A. S. Krayev, D. A. Kramerov, K. G. Skryabin, A. P. Ryskov, A. A. Baev, G. P. Georgiev, *Nucleic Acids Res.* **8**, 1201 (1980).
22. S. R. Haynes, T. P. Toomey, L. Leinwand, W. R. Jelinek, *Mol. Cell. Biol.* **1**, 573 (1981).
23. Supported by NIH grant GM 06965. K.W.M. was initially supported by a postdoctoral fellowship from the National Institutes of Health. B.T.S. was supported by a predoctoral fellowship from the National Science Foundation. Y.H.S. and K.A.E. were supported by a national research service award (1 T32 GM07616) from the National Institute of General Medical Sciences. We thank M. Douglas for computer graphics, T. Hunkapiller for computer programs, and B. Larsh for manuscript preparation. We thank W. R. Jelinek for pointing out the presence of several Alu-like sequences in gene 27.1.

* To whom correspondence should be addressed.

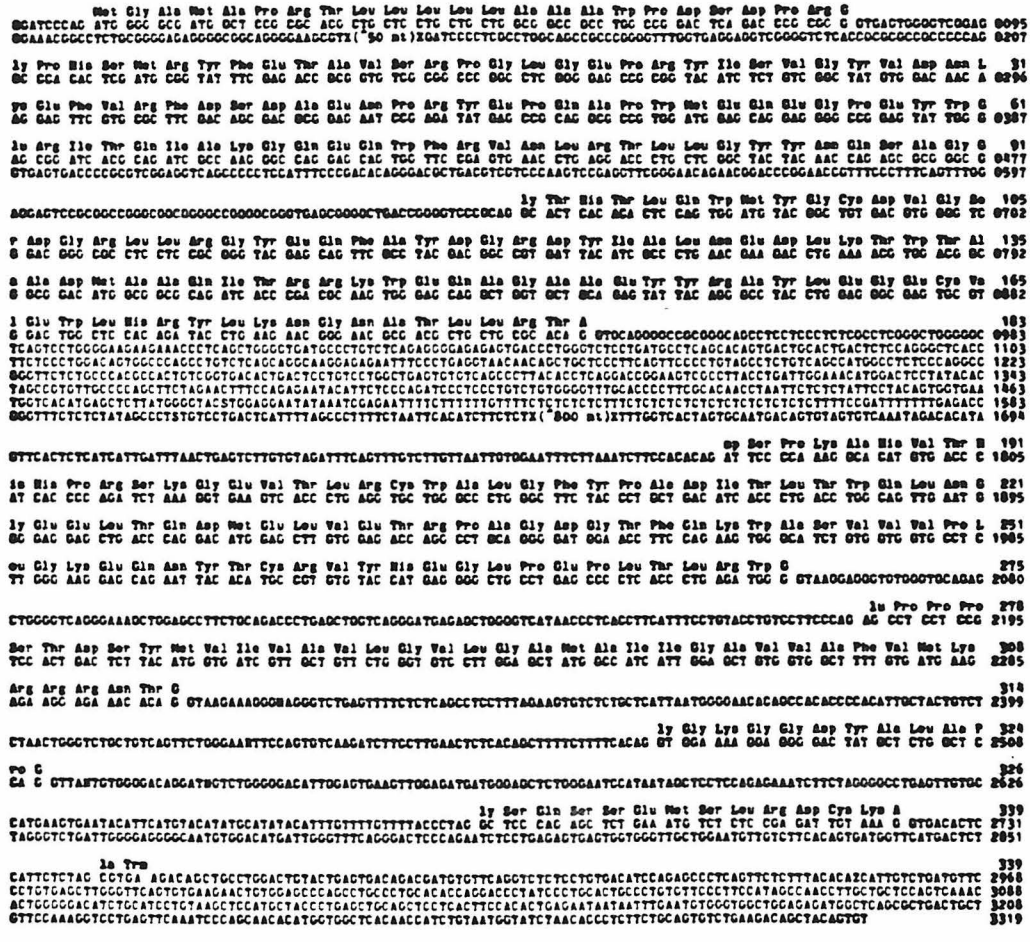
5 November 1981; revised 16 December 1981

The following is the corrected sequence of the H-2L^{d} gene. As originally published, the sequence was in error due to a cloning artifact. In addition, a number of editorial mistakes were made. The cloning artifact affected the first 90 nucleotides of the sequence. It resulted from a Bam/Rsa-filled in ends ligation, and the sequence at the joint reflects the presence of some M13 sequence as well. This artifact also affected the restriction map, as the affected clone was used as a probe for exon I sequences.

The cloning artifact and most of the editorial mistakes were corrected by extensive re-reading of the original sequencing gels and by resequencing of the clone indicated in the diagram below. In addition, Henry Sun has checked the sequence of the second intron and third exon by making new M13 subclones and sequencing them. The nucleotides affected by the mistakes and subsequently corrected are underlined in the copy of the original sequence diagram that follows the corrected sequence. There were seven mistakes that resulted in codon changes and 32 other changes in addition to the cloning artifact. The restriction map that follows the sequence figures shows the sites that were affected by the cloning artifact.

An unrelated mistake lay in the misidentification of the location of the mouse B1 and B2 repetitive sequences(called Alu-like sequences in the paper). Improved computer software has allowed the precise identification of their locations. The B1 sequence in the 27.1 pseudogene occupies nucleotides 2580-2660;

the B2 sequence is located at nucleotides 3727-3893, just before the beginning of exon IV. The H-2L^d sequence also contains a B1 sequence which begins at nucleotide 1587 and continues up to the gap in the large intron. This B1 sequence is in a postion analogous to that of the B1 sequence in the third intron of the 27.1 pseudogene. The H-2L^d sequence lacks a B2 sequence, however. The region of nonhomology between the H-2L^d and 27.1 sequences that shows up in the third intron corresponds to the T-rich area that precedes the B1 sequence: the length and sequence of this region is variable in the two sequences.



EXON 1

Met Ala Pro Arg Thr Leu Leu Leu Leu Leu
CCAGGGCGGATTGAGAGGGAAGACCACACACCCTGTGAGCTCACTGTGTCCAGTGAGTAGCTGCACTGGGGTCCACAGCTCACTCCAGGGATCACTCCCAGATGGGGGCG ATG GCT CCG CGC ACG CTG CTC CTG CTG CTG 140

Ala Ala Ala Trp Pro Asp Ser Asp Pro Arg G
GCG GCC GCC TGG CCC GAC TCA GAC CCG CGG G GTGAGTGGGGGTGCGGAGGGAAACGGCCTCTGCGGGGAGAGGGGCGGCAGGGGGAAGCGC ...GATCCCCTCGCCTCGCAGCCGCCCGGGGTTTGGTAGGAGGTCGG 278

last codon
EXON 2

ly Pro His Ser Met Arg Tyr Phe Glu Thr Val Ser Arg Arg Gly Leu Gly Glu Pro Arg Tyr Ile Ser Val Gly Tyr Val Asp Asn Lys Glu Phe
GGTCTCACCGCGCGCCCGCCCCAG GC CCA CAC TCG ATG CGG TAT TTC GAG ACC GTG TCC CGG CGC GGC CTC GGG GAG CCC CGG TAC ATC TCT GTC GGC TAT GTG GAC AAC AAG GAG TTC 387

Val Arg Phe Asp Ser Asp Ala Glu Asn Pro Arg Tyr Glu Pro Arg Ala Pro Trp Met Glu Gln Glu Gly Pro Glu Tyr Trp Glu Arg Ile Thr Gln Ile Ala Lys Gly Gln Glu
GTG CGC TTC GAC AGC GAC GCG GAG AAT CCG AGA TAT GAG CCG AGC GCG CCG TGG ATG GAG CAG GAG GGG CCG GAG TAT TGG GAG CGG ATC ACG CAG ATC GCC AAG GGC CAG GAG 511

Gln Trp Phe Arg Val Asn Leu Arg Thr Leu Leu Gly Tyr Tyr Asn Gln Ser Ala Gly G
CAG TGG TTC CGA GTG AAC CTG AGG ACC CTG CTC GGC TAC TAC AAC CAG AGC GCG GGC G GTGAGTGACCCCGCGTCGGAGGTCACGGCCCCCTCCCTTTCCCGACACAGGGACGCTGACTTCGTACCCAAGT 644

ly Thr His Thr Leu Gln Trp Met Tyr
CCGAAGTTCGGGAACAGAACGGACCCGGAACCAGTTTCCCTTTCAGTTTGGAGGAGTCGCGGCCGGGCGGCGGGGCCGGGGCGGGTGAGCGGGGCTGAACCGGGGTACCGCAG GC ACT CAC ACA CTC CAG TGG ATG TAC 780

EXON 3

Gly Cys Asp Val Gly Ser Asp Gly Arg Leu Leu Arg Gly Tyr Glu Gln Phe Ala Tyr Asp Gly Cys Asp Tyr Ile Ala Leu Asn Glu Asp Leu Lys Thr Trp Thr Phe Ala Asp
GGC TGT GAC GTG GGG TCG GAC GGG CGC CTC CTC CGC GGG TAC GAG CAG TTC GCC TAC GAC GGC TGC GAT TAC ATC GCC CTG AAC GAA GAC CTG AAA ACG TGG ACG TTC GCG GAC 894

Met Ser Ser Gln Ile Thr Arg Arg Lys Trp Glu Gln Ala Gly Ala Ala Glu Tyr Tyr Arg Ala Tyr Leu Glu Gly Glu Cys Val Glu Trp Leu His Arg Tyr Leu Lys Asn Gly
ATG TCG TCG CAG ATC ACC CGA CGC AAG TGG GAG CAG GCT GGT GCT GCA GAG TAT TAC AGG GCC TAC CTG GAG GGC GAG TGC GTG GAG TGG CTC CAC AGA TAC CTG AAG AAC GGG 1008

Asn Ala Thr Leu Leu Arg Thr A
AAT GCT ACG CTG CTG CGC ACA G GTGCAGGGGCCGCGGCGGGCAGCTCCTCCCTCTCGCCTCGGGCTGGGGGCTCAGTCCTGGGGAAGAAGAAACCCTCAGCTGGGGTGATGCCCTGTCTCAGAGGGGAGAGAGTGACCCTG 1150

GGTTCCTGATGCCTCAGCACAGTGACTGCACTGACTCTCCCAGGGCTCAGCTTCTCCCTGGACAGTCCCAGCCTGTCTCAGGAGGGAAGGAGAGAAATTTCCCTGAGGTAACAACAGCTGCTCCCTTCAGTTCCCCTGTAGCCTCTGTCAG 1300

CCATGGCCTCTCCCAGGCGGGTTCTCTGCCCACGCCACTGTCGGTGACACTGACTCCTGTCCTGGCTGATGTGTGTCAGCGCCTTACACCTCAGGACCGGAAGTCGCCTTACCTGATTGGAAACATGGACTCCTATACACTAGCCGTGTT 1450

GCCCCAGCTTCTAGAACTTTCCAGAGAATACATTCTCCCAGATCCCTCCCTGTCTGTGGGGTTTGCACCCCTTCGCACACAACCTAATTCTCTCTATTCCTACAGTGGTGAATGGTCACATGAGCTCTTATGGGGTACSTGGAGGAATATAA 1600

ATCGAGAATTTTCTTTTTTGTTTTCTCTCTCTCTTTCTCTCTCTCTCTCTCTCTCTCTTTTCCGATTTTTTTGAGACCGGGTTTCTCTCTATAGCCCTGTGTCCTGACTCATTTTAGCCCTTTTCTAATTCACATCTTCTCT ...TT 1749

ep Ser Pro Lys Ala
TGGTCACTAGTGCAATGACAGTGTAGTGTCAAATAGACACATAGTTCACTCTCATCATTGATTTAACTCAGTCTTGTGTAGATTTCAGTTTGTCTTGTTAATTGTGGAATTTCTTAAATCTTCCACACAG AT TCC CCA AAG GCA 1883

EXON 4

His Val Thr His His Pro Arg Ser Lys Gly Glu Val Thr Leu Arg Cys Trp Ala Leu Gly Phe Tyr Pro Ala Asp Ile Thr Leu Thr Trp Gln Leu Asn Gly Glu Glu Leu Thr
CAT GTG ACC CAT CAC CCC AGA TCT AAA GGT GAA GTC ACC CTG AGG TGC TGG GCC CTG GGC TTC TAC CCT GCT GAC ATC ACC CTG ACC TGG CAG TTG AAT GGG GAG GAG CTG ACC 2007

Gln Asp Met Glu Leu Val Glu Thr Arg Pro Ala Gly Asp Gly Thr Phe Gln Lys Trp Ala Ser Val Val Val Pro Leu Gly Lys Glu Gln Asn Tyr Thr Cys Arg Val Tyr His
CAG GAC ATG GAG CTT GTG GAG ACC AGG CCT GCA GGG GAT GGA ACC TTC CAG AAG TGG GCA TCT GTG GTG GTG CCT CTT GGG AAG GAG CAG AAT TAC ACA TGC CGT GTG TAC CAT 2121

Glu Gly Leu Pro Glu Pro Leu Thr Leu Arg Trp G
GAG GGG CTG CCT GAG CCC CTC ACC CTG AGA TGG G GTAAGGAGGGTGTGGGTGCAGAGCTGGGGTCAGGGAAAGCTGGAGCCTTCTGCAGACCCTGAGCTGGTCAGGGATGAGAGACTGGGGTCATAACCCTCACCTTC 2250

EXON 5

lu Pro Pro Pro Ser Thr Asp Ser Tyr Met Val Ile Val Ala Val Leu Gly Val Leu Gly Ala Met Ala Ile Ile Gly Ala Val Val Ala Phe Val
ATTTCCTGTACCTGTCCTTCCCAG AG CCT CCT CCG TCC ACT GAC TCT TAC ATG GTG ATC GTT GCT GTT CTG GGT GTC CTT GGA GCT ATG GCC ATC ATT GGA GCT GTG GTG GCT TTT GTG 2378

Met Lys Arg Arg Arg Asn Thr G
ATG AAG AGA AGG AGA AAC ACA G GTAAGAAAGGGHAGGGTCTGAGTTTTCTCTCAGCCTCCTTTAGAAGTGTCTCTGCTCATTAATGGGGAACACAGCCACACCCCACATTGCTACTGTCTCTAACTGGGTCTGCTGTCAGTT 2520

EXON 6

ly Gly Lys Gly Gly Asp Tyr Ala Leu Ala Pro G
CTGGGAARTTCCAGTGTCAAGATCTTCCTTGAACTCTCACAGCTTTXCTTTTCACAG GT GGA AAA GGA GGG GAC TAT GCT CTG GCT CCA G GTTAATGTGGGGACAGGATWGTTCTGGGGGACATTGGAGTGAAGTTG 2657

EXON 7

ly Ser Gln Ser Ser Glu
GAGATGATGGGAGCTCTGGGAATCCATAATAGCTCCTCCAGAGAAATCTTCTAGGGGCCTGAGTTGTGCCATGAAGTGAATACATTCATGTACATATGCATATACATTTGTTTTGTTTTACCCTAG GC TCC CAG AGC TCT GAA 2800

Met Ser Leu Arg Asp Cys Lys A
ATG TCT CTC CGA GAT TGT AAA G GTGACACTCTAGGGTCTGATTGGGGAGGGGCAATGTGGACATGATTGGGTTTCAGGGACTCCCAGAATCTCCTGAGAGTGAGTGGTGGGTTGCTGGAATGTTGTCTTCACAGTGATGGTT 2942

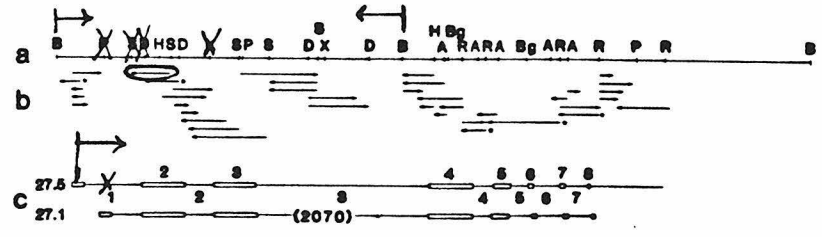
EXON 8

le Trp
CATGACTCTCATTCTCTAG CG TGA AGACAGCTGCCTGGACTGTACTGAGTGACAGACGATGTGTTCAGGTCTCTCCTGTGACATCCAGAGCCCTCAGTTCTCTTTACACAZCATTGTCTGATGTTCCCTGTGAGCTTGGGTTCAGTG 3088

TGAAGAACTGTGGAGCCCAGCCTGCCCTGCACACCAGGACCCTATCCCTGCACTGCCCTGTGTTCCCTTCCATAGCCAACCTTGCTGCTCCAGTCAAACACTGGGGGACATCTGCATCCTGTAAGCTCCATGCTACCCTGAGCTGCAGCT 3238

CCTCACTTCCACACTGAGAATAATAATTTGAATGTGGGTGGCTGGAGAGATGGCTCAGCGCTGACTGCTCTTCCAAAAGGTCCTGAGTTCAAATCCCAGCAACACATGGTGGCTCACAACCATCTGTAATGGTATCTAACACCCTCTTCTG 3388

3414



Appendix C

Identification of Class I Genes.

This article was published in Nature.

Identification of the class I genes of the mouse major histocompatibility complex by DNA-mediated gene transfer

Robert S. Goodenow*, Minnie McMillan†, Margery Nicolson‡, Beverly Taylor Sher*, Kurt Eakle*, Norman Davidson§ & Leroy Hood*

* Division of Biology, California Institute of Technology, Pasadena, California 91125, USA
† Department of Microbiology, USC Medical School, Los Angeles, California 90033, USA
‡ AMGen, Newbury Park, California 91320, USA
§ Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA

DNA-mediated gene transfer was used to identify cloned class I genes from the major histocompatibility complex of the BALB/c mouse. Three genes encoding the transplantation antigens H-2 K^d, D^d and L^d were identified as well as genes encoding the Qa-2,3 and two TL differentiation antigens. As many as 10 putative novel class I genes were detected by the association of their gene products with β_2-microglobulin. Alloantiserum prepared to one of the novel antigens was used to demonstrate the expression of the previously undetected antigen on spleen cells of various inbred, congeneic, and recombinant congeneic strains of mice.

THE major histocompatibility complex of the mouse encodes a family of cell-surface antigens known as the class I molecules. These gene products, which fall into two categories, are encoded by distinct regions of the major histocompatibility complex. The H-2 region encodes the transplantation antigens which were initially defined by allograft rejection[1,2]. These antigens are expressed on most somatic cells of the mouse and are involved in the T-cell recognition of cells altered by viral or neoplastic transformation (for a review see refs 3, 4). The second category of class I antigens are the haematopoietic differentiation antigens encoded by genes in the Tla region. The Tla region encodes three distinct types of differentiation antigens, Qa-1, Qa-2,3 and TL[5]. The TL antigens are expressed only on certain subpopulations of thymocytes and their neoplastic counterparts in certain T-cell leukaemias[6]. The Qa-1 antigen is expressed by thymocytes, peripheral T and B cells, as well as lymphoblasts[7]. The Qa-2,3 antigens are found on a subset of bone marrow-derived cells[8]. The biological functions of the TL or Qa antigens are not known.

Class I antigens have been characterized using specific alloantisera and monoclonal antibodies prepared by the appropriate cross-immunizations with cells from inbred, congeneic and recombinant strains of mice. Serological analysis has demonstrated that the transplantation antigens are extremely polymorphic and that different strains of mice have distinct combinations or haplotypes of class I alleles. Particular class I molecules are denoted by appropriate letters with a superscript small letter for the haplotype. For example, the serologically defined transplantation antigens of the BALB/c mouse which is of the $H\text{-}2^d$ haplotype are denoted K^d, D^d, L^d or R^d.

Class I molecules of both categories appear to share a common structure. The class I polypeptide is a 45,000-molecular weight integral membrane protein which is noncovalently associated with a 12,000-molecular weight polypeptide, β_2-microglobulin. Class I molecules consist of three external domains, each of ~90 amino acids, a transmembrane region of ~40 residues, and one cytoplasmic domain of 30 residues[9,10].

Class I cDNA clones have been characterized and used as probes to obtain genomic clones containing class I genes[11-17]. Two genes isolated from a BALB/c Crgl sperm DNA library constructed with the lambda bacteriophage (λ) vector have been extensively characterized. The class I gene from clone λ27.1 is a pseudogene which has been mapped to the Qa-2,3 subregion of the Tla region by restriction enzyme site polymorphisms[15]. DNA sequence analysis of gene 27.1 shows that this class I gene is divided into eight distinct exons encoding each of the individual domains of a class I molecule. Recently, DNA-mediated gene transfer and radioimmunoassays were used to demonstrate that a gene from clone λ27.5 encodes the $H\text{-}2L^d$ transplantation molecule[18]. The L^d molecules produced by mouse L cells transformed with λ27.5 DNA were shown to be virtually indistinguishable from the L^d molecules expressed on spleen cells from BALB/c mice by biochemical, immunological and functional criteria[18-21]. The DNA sequence of gene 27.5 is identical to that of the L^d antigen at all 77 positions that can be compared[22]. Other laboratories have also used gene

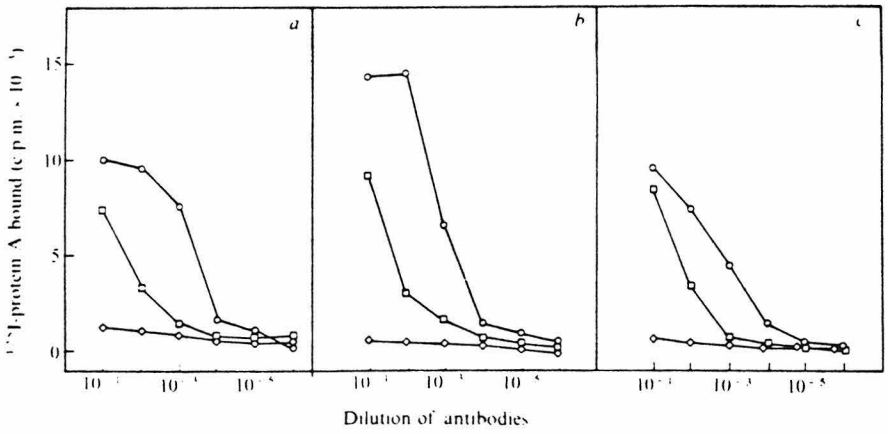


Fig. 1 Radioimmunoassay of λ1.3, λ21.1 and λ26.1 transformants. Transformants which reacted in the panel analysis (Table 1) with antibodies 34-1-2 to K^d molecules were tested by the cell-binding radioimmunoassay with: ○, 34-1-2 antibodies to K^d; □, 20-8-4 antibodies to $K^{b\,42}$ which react with K^d molecules; and ◇, 34-2-12 antibodies to D^d. Transformants derived from transfection with DNA from clones λ26.1 (*a*), λ21.1 (*b*) and λ1.3 (*c*). Spleen and Ltk^- controls are shown in Fig. 4.

Table 1 Summary of the coding assignments for the class I genomic clones

Clone	Reaction of transformants with antibodies specific for K^d	D^d	L^d	Qa-2,3	TL^c	Assignment
λ1.3	+					K^d
λ26.1	+					K^d
λ21.1	+					K^d
c17.1	+					K^d
c18.1		+				D^d
λ27.5			+			L^d
c59.2			+			L^d
c50.2				+		Qa-2,3
λ17.3					+	TL
c6.3					+	TL
λ24.8					+	TL
66.1					+	TL

Mouse L tk^- cells were co-transformed with ~1 μg of DNA from each of the λ and cosmid clones (see text) and the HSV tk gene in pBR322 (ptk5) as previously described[18]. For those cosmid clones containing more than one gene, the clone DNA was digested with an appropriate restriction enzyme to inactivate all but a single gene for transformation (see ref. 24). Uncloned tk^- transformants selected in hypoxanthine/aminopterin/thymidine medium[38] were panel analysed by radioimmunoassay as previously described[18] using monoclonal hybridoma antibodies 34-1-2 to K^d (ref. 39), 34-2-12 to D^d (ref. 39), 30-5-7 to L^d (ref. 40), D3.262 to Qa-2,3 (ref. 41) and TLm(i) to TL[33]. Positives are indicated for those transformants which yielded at least 5,000 c.p.m. of iodinated (^{125}I) protein A bound (see Figs 2, 4) with a 10^{-2} or 10^{-3} dilution of the ascites fluid in duplicate transformation experiments. Background for the various antibodies and transformants ranged over 100 to 300 c.p.m., except for the IgM Qa antibodies which utilized a facilitating reagent (see Fig. 4). The active genes on cosmid clone 17.1, 18.1, 59.2, 50.2, 6.3 and 66.2 are the first gene on cluster 11, the first gene on cluster 13, the second gene on cluster 2, the second gene on cluster 6, the second gene on cluster 4, and the first gene on cluster 5, respectively (see Fig. 5).

transfer to identify the cloned class I genes from libraries constructed from tissues or tumour cell lines from other sub-strains[17] and strains[23] of mice.

In an attempt to assess the total number of class I genes contained in the BALB/c mouse, genomic clones containing class I genes have been isolated from a cosmid library constructed from BALB/c sperm DNA[24]. Fifty-four cosmid clones containing 36 unique class I genes have been assigned to 13 distinct groups or clusters encompassing 837 kilobases of DNA. These include most of the class I genes in the BALB/c mouse.

Here, we have used gene transfer and serological analyses to examine individual class I genomic clones from the λ and cosmid libraries. Cloned class I genes encoding the serologically defined H-2 K^d, D^d, L^d, Qa-2,3 and TL antigens have been identified. Evidence is also presented for as many as 10 additional novel class I genes encoding products not previously defined serologically.

Analysis of 96 λ and 36 cosmid class I clones by gene transfer

In an attempt to examine most class I genes from the BALB/c mouse, 96 genomic clones isolated from the λ library and 36 genomic clones previously isolated from the cosmid library (designated c) were analysed. DNA from each of these genomic clones, together with the herpes viral thymidine kinase gene, was used to co-transform thymidine kinase-negative (tk^-) mouse L cells as previously described[18]. The class I H-2^d products of the transformed L cells were detected by radioimmunoassay and analysed by two-dimensional gel electrophoresis. Since mouse L cells are fibroblasts derived from C3H mice of the *H-2*^{k} haplotype, the endogenous K^k and D^k transplantation antigens can readily be distinguished from any of the BALB/c class I molecules of the *H-2*^{d} haplotype by appropriate monoclonal antibodies. The mouse L-cell fibroblasts do not express the differentiation antigens Qa-1, Qa-2,3

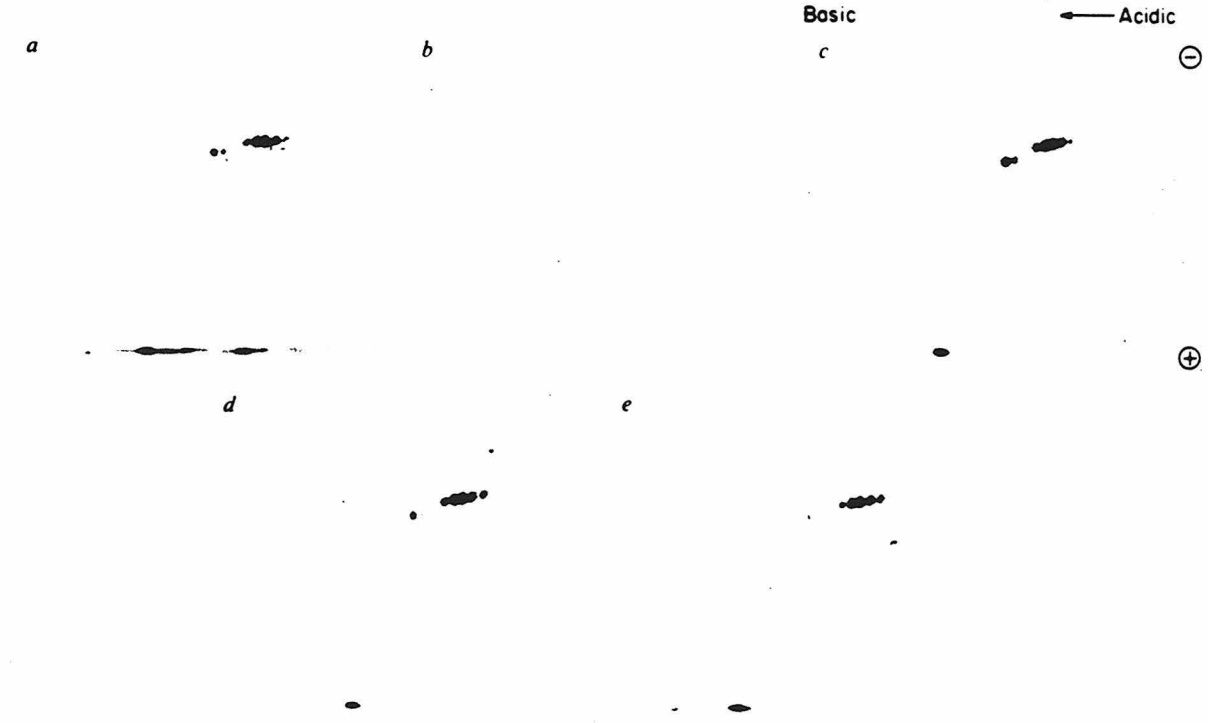


Fig. 2 Fluorographs of two-dimensional gels of H-$2K^d$ molecules immunoprecipitated from cell lysates of BALB/cJ spleen lymphocytes (*a*), tk^- cells transformed only with the tk gene (*b*), λ1.3 (*c*), λ26.1 (*d*) and λ21.1 (*e*) transformants. Cells were biosynthetically radiolabelled, immunoprecipitated with 20-8-4 monoclonal antibodies (supernatant fluids), and analysed by two-dimensional gel electrophoresis as previously described[18]. The gels were prepared for fluorography, dried and exposed for ~7 days.

or TL. These results are summarized in Table 1. In almost all cases, the binding of specific antibodies to the transformants was comparable to that of the binding of the antibodies to the corresponding antigens and not cross-reacting antigens (see Fig. 4).

Identification of genes encoding the H-2 K, D and L transplantation antigens

The L-cell transformants derived from transfections with the λ1.3, λ26.1 and λ21.1 clones express K^d molecules which are virtually indistinguishable in their patterns of serological reactivity with two different monoclonal antibodies to the K^d molecule (Fig. 1), both of which react as strongly with these transformants as they do with spleen cells (see Fig. 4). No reactivity is observed with these four transformants using monoclonal antibodies specific for D^d or L^d antigens (see Fig. 1 legend). Antibodies specific for the K^d molecule do not react with mouse L cells transformed only with the thymidine kinase gene (see Fig. 4). These are designated tk^- transformants. To determine whether these three class I genes encode distinct gene products, the antigens were isolated by immunoprecipitation and analysed by two-dimensional gel electrophoresis (Fig. 2). As the class I antigens of different haplotypes each have distinct constellations of spots[25], we believe this approach is very discriminating. The products of class I genes 1.3, 26.1 and 21.1 appear extremely similar to one another and to the K^d molecules isolated from normal BALB/c spleen lymphocytes. However, minor electrophoretic differences were observed in the faint, lower molecular weight species (Fig. 2*a* and 2*e*).

As the electrophoretic patterns between products obtained from duplicate transformations appear identical, the variation in the lower molecular weight products may reflect differences in post-translational modifications of these molecules or differential alternative splicing, as has been previously suggested for class I genes[15]. Restriction map analyses of clones λ1.3, λ26.1 and λ21.1 (Fig. 3), however, suggest that the K^d genes they contain are identical to one another and also to the first gene on cluster 11 or clone c17.1 (see Fig. 5) which contains a K^d gene by transformation (Table 1, Fig. 4*a*). Thus, all of these clones appear to encode very similar or identical genes on an independently derived series of what we believe are overlapping clones, although we cannot exclude the possibility that certain of these genes may actually be distinct and yet appear identical by the limited restriction mapping we have carried out to date. These genes are currently being sequenced in order to resolve this issue.

Cosmid clone 18.1 contains an H-2D^d gene as demonstrated by the radioimmunoassay of the c18.1 transformants with monoclonal antibodies to theD^d antigen (Table 1, Fig. 4*b*). The 18.1 transformants reacted with antibodies specific for D^d products and not reagents which detect K^d, L^d, TL or Qa antigens. The D^d molecules expressed by the transformants are currently being compared with the molecules isolated from spleen cells by two-dimensional gels and tryptic peptide map analysis to confirm this identification. No lambda clones containing an H-2 D^d gene were identified (see below).

An H-2L^d gene has been identified on cosmid clone 59.2 by transformation (Table 1) and comparison of restriction maps between the c59.2 and λ27.5 genes[24]. The L^d molecules expressed on the corresponding transformants appear indistinguishable in radioimmunoassays from the products of the previously identified gene contained in clone λ27.5 (Fig. 4*c*).

The analyses based on 132 independent transformations resulted in the identification of at least one H-2K^d, a single H-2D^d and one H-2L^d gene. Serological evidence suggests that in some mice there may be multiple K and D gene products[26–31], and additional transplantation antigens such as R^d (ref. 32). The failure to find multiple class I K^d and D^d genes may be explained in several ways. First, the serological multiplicity of K or D antigens may arise from post-translational modifications or alternative patterns of RNA splicing. Second, different inbred strains may have different numbers of K and D genes. Third, multiple class I genes might appear identical by restriction mapping and still differ in their DNA sequences. In this case, multiple distinct genes might be confused with a single gene present on overlapping clones. Finally, these genes may not have been detected for various technical reasons, for example, failure to clone the gene or express the gene at levels detectable in the radioimmunoassay. For example, the radioimmunoassay is less effective with heterogeneous alloantisera and specific monoclonal reagents for the R^d or the Qa-1 gene products are lacking.

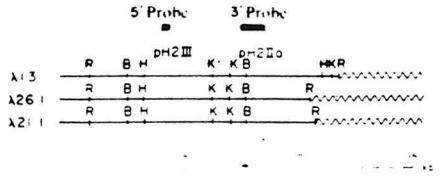


Fig. 3 Restriction maps of the lambda clones containing the K^d genes. DNA from each of the λ clones which yielded K^d-positive transformants was restriction-mapped with the indicated enzymes by single and double digestion as previously described[24]. The 5' probe pH-2III and the 3' probe pH-2IIa[11] were used to detect fragments containing coding sequences homologous to the class I genes (dashed lines). The pH-2III probe contains the sequences encoding amino acids 63–160 of a transplantation antigen and pH-2IIa amino acids 167–352. The coding sequences were orientated using sequencing subclones as previously described[22]. The solid line indicates the insert DNA, and the λ vector DNA is designated by the wavy lines. R, *Eco*RI; B, *Bam*HI; H, *Hin*dIII; K, *Kpn*I. The 3' *Eco*RI sites mark the linkers used in constructing the library.

Fibroblast expression of gene encoding the mouse haematopoietic differentiation antigen Qa-2,3

Transformants from c50.2 DNA seem to express low levels of the Qa-2,3 antigen as detected by radioimmunoassay with monoclonal antibodies specific for these products (Fig. 4*d*). The BALB/c substrain from which the library was derived is Qa-2,3-positive (L. Flaherty, personal communication). It is interesting that *Qa-2,3*, a gene normally expressed only in haematopoietic cells, appears to be expressed in transformed fibroblasts. The fact that the Qa-2,3 differentiation antigen is expressed at low levels on haematopoietic cells[8] may explain why limited numbers of Qa-2,3 molecules are expressed in the transformed mouse L cells. It is difficult to ascertain the basis for the low level of expression of this antigen by the transformants. Several possibilities are now being investigated. For example, only rare transformants may have incorporated sufficient copies of the gene to override differentiation signals. In addition, glycosylation of the polypeptide in fibroblasts might, in this particular instance, alter the antigenicity of the molecule causing a less avid reaction with the antibodies. The fact that the c50.2 transformants exhibit elevated levels of β_2-microglobulin (see below) does suggest that significant amounts of the antigen may be expressed in an altered form associated with β_2-microglobulin. Accordingly, it has been difficult to obtain sufficient Qa-2,3 antigen from spleen cells or transformed L cells for two-dimensional gel analysis. The identification of a Qa-2,3 gene should readily permit its characterization at the DNA level.

Two different class I genes encode two distinct TL T-cell differentiation antigens

The TL antigens are expressed only on thymocytes at a certain

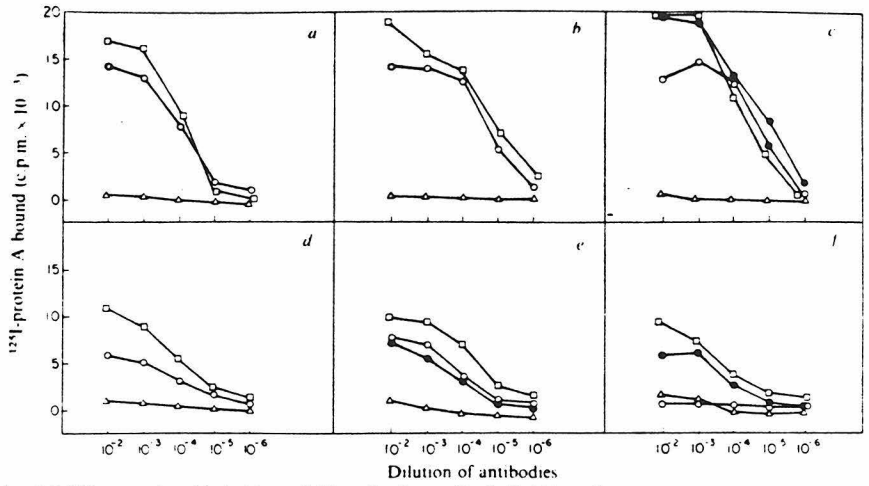


Fig. 4 Radioimmunoassay of the transformants expressing K^d, D^d, L^d, Qa-2,3 and TL antigens. *a*, Transformants derived from transfection of L tk^- cells with DNA from c17.1 (○) were tested by radioimmunoassay for the expression of K^d molecules using 34-1-2 antibodies to K^d. *b*, Transformants from transfection with DNA from c18.1 (○) and shown to express D^d molecules as detected in the panel analysis were analysed using 34-2-12 antibodies to D^d. *c*, Transformants derived from transfection with DNA from λ27.5 (○), which contains the L^d gene[18,22], and transformants from DNA from c59.2 (●) were tested against 30-5-7 antibodies to L^d. *d*, Transformants from transfection with c50.2 DNA (○) were tested in a cell-binding radioimmunoassay using D3.262 monoclonal hybridoma IgM antibodies to Qa-2. Rabbit antiserum to mouse IgM (Bionetics) was used as a second step reagent before incubation of the cells with ^{125}I-protein A. *e*, Transformants derived from transfection with DNA from λ17.3 were tested against antibodies to TL (●) TLm(i) and (○) TLm(iii) (see text). *f*, Transformants from transfection with λ24.8 DNA were tested against the same antibodies to TL indicated in *e*. BALB/c spleen cells, panels *a–d* (□) or thymocytes, panels *e–f* (□) and Ltk^- cells, panels *a–f* (△) were run as controls. Thymocytes and Ltk^- cells in panel *e* were tested against TLm(i), panel *f* against TLm(iii).

stage of differentiation. In some inbred strains of mice, distinct TL molecules are expressed on certain leukaemic cells of T-cell origin[6]. Two monoclonal antibodies can distinguish these TL antigens. Antibodies denoted TLm(i) recognize TL determinants present on both normal and leukaemic cells, whereas antibodies designated TLm(iii) react only with the TL determinants on normal thymocytes[33]. Figure 4*e*, *f* shows that transformants from λ clone 17.3 reacted with both monoclonal antibodies at levels comparable to those observed with BALB/c thymocytes (Fig. 4*e*, *f*), whereas transformants from λ clone 24.8 reacted only with monoclonal antibodies TLm(i). These observations suggest that there are at least two distinct genes, λ17.3 and λ24.8, which encode two TL antigens. From the serological analyses described above, gene λ17.3 may be expressed on normal BALB/c thymocytes and leukaemic cells whereas gene λ24.8 would be expressed only on leukaemic cells. The λ17.3 and λ24.8 gene products are currently being characterized and compared with the molecules present on normal and leukaemic thymocytes by two-dimensional gels to clarify this point further. The cosmid clones containing the corresponding TL genes have been identified by transformation (Table 1, Fig. 5). The 17.3 TL gene appears to be the first gene on cosmid cluster 5 or clone 6.3, and the 24.8 gene the second gene on cosmid cluster 4 or clone 66.1. Since each TL gene maps to a separate cosmid cluster, they must be distinct genes. These results illustrate two important points regarding gene transfer into mouse L cells. First, as was noted for the Qa-2,3 gene, the TL genes which are developmentally regulated and expressed only on thymocytes can be expressed in fibroblasts as a result of transformation. Second, genes not expressed in the mouse except on leukaemias, such as gene λ24.8, can be expressed by transformed fibroblasts. It seems unlikely that the activation of endogenous L-cell genes encoding similar TL molecules would account for these results as restriction diges-

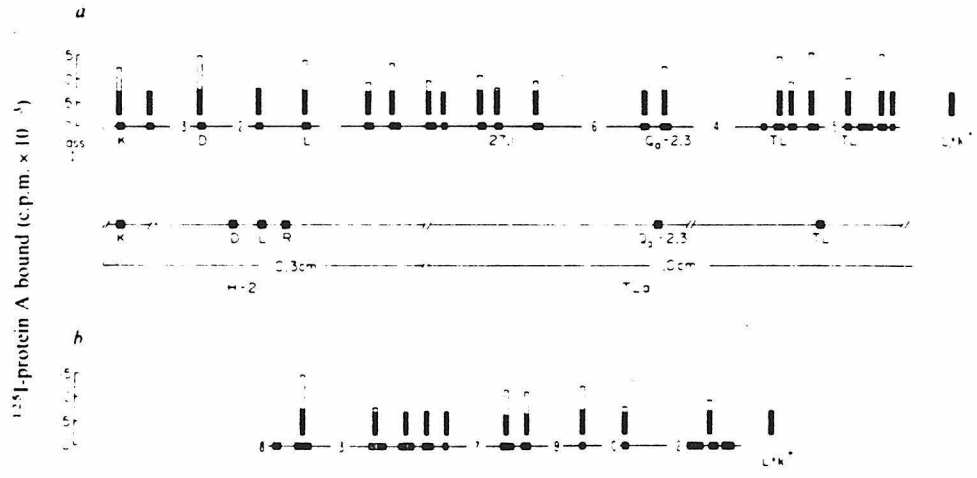


Fig. 5 Radioimmunoassay of cosmid transformants for β_2-microglobulin. The mixed population of tk^- transformants derived from transfection with a single gene from each of the various cosmid clones, obtained by restriction digestion, was tested for the expression of β_2-microglobulin in a cell-binding radioimmunoassay using rabbit antiserum to rat β_2-microglobulin. Each transformant was assayed in duplicate against a 10^{-2} dilution of the antiserum. The numbers shown at the left indicate the c.p.m. of ^{125}I-protein A bound per 50,000 cells. Less than 10% variation in the number of c.p.m. was observed between duplicate transformations or assays. The dark portion of each bar represents the basal level of L-cell β_2-microglobulin measured for tk^- transformants which is the average of duplicate samples tested (right). The open portion of each bar represents the total number of c.p.m. of ^{125}I-protein A bound for each transformant which exceeds the level on tk^- cells. Cosmid clusters are represented schematically with the genes indicated as the enclosed boxes. Each cluster is comprised of a group of noncontiguous overlapping clones[24]. The cosmid cluster number is indicated to the left of the horizontal black line and the genes encoding the serologically defined products are indicated below the appropriate boxes. *a*, β_2-microglobulin on the surface of transformants derived from transfection with DNA from cosmid clones linked with mapped genes or genes encoding serologically defined antigens; *b*, undefined clusters only.

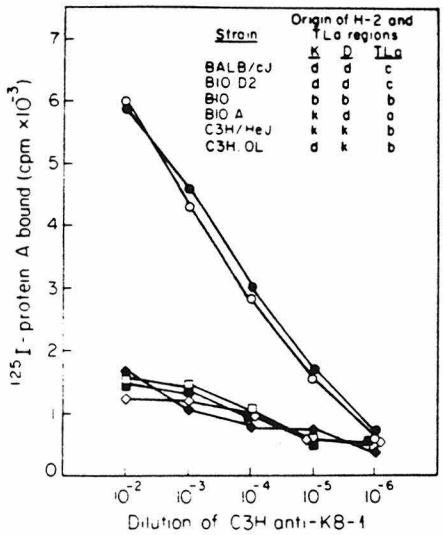


Fig. 6 Radioimmunoassay of spleen cells with C3H antibodies to K8-1 transformants. Four C3H mice were immunized by injecting 5×10^7 UV-irradiated K8-1 cells intraperitoneally every 2 weeks for 8 weeks. Mice were bled 7 days after the last injection and the antiserum used in the cell-binding radioimmunoassay against spleen cells as previously described[18]. Reaction of spleen cells from: ●, BALB/cJ; ○, B10.D2; ■, C3H/HeJ; □, C3H.OL; ◆, B10; ◇, B10.A. The origins of the H-2 and Tla regions for each of the mice are indicated in the panel.

tion of λ17.3 or λ24.8 DNA destroys the ability of these genes to transform L cells to the TL-positive phenotype.

Ten class I genes encode serologically undefined gene products

Of the 36 distinct class I clones, six contain genes encoding serologically defined gene products (Fig. 5*a*). The remaining class I genes are either pseudogenes or encode as yet unidentified cell-surface products. For example, they could encode nonpolymorphic molecules expressed on the surface of most cells which therefore would not have been detected by classical serological approaches. Alternatively, they might be differentiation antigens expressed at defined stages on tissues not yet tested for their ability to generate alloantisera. Cytoplasmic or secreted forms of class I-molecules might represent another category of uncharacterized products.

We have used a simple but effective approach to identify class I gene products for which there are no specific serological reagents. All class I molecules studied associate with β_2-microglobulin before expression on the cell surface[34,35]. Therefore, the expression of foreign class I molecules on the surface of the mouse L cell should result in the elevation of the amount of β_2-microglobulin expressed on the cell surface. To test this hypothesis, a radioimmunoassay for cell-surface β_2-microglobulin was developed and used to quantitate the amount of β_2-microglobulin on several transformants expressing foreign serologically defined class I molecules as well as several which failed to react with any of the class I-specific antibodies. The tk$^-$ transformants express a basal level of β_2-microglobulin as do the mixed population of tk$^-$ transformants derived from transfection with the L^d gene cleaved into small coding fragments with *Bam*HI enzyme. In contrast, mouse L cells transformed with the L^d, D^d, K^d and TL genes all expressed levels of cell-surface β_2-microglobulin well above the background level as measured by radioimmunoassay (Fig. 5*a*). This increase is reproducible in that the measured levels of β_2-microglobulin obtained for duplicate assays or transformations vary by less than 10%. In addition, the elevation in β_2-microglobulin expression appears to be a stable phenotype, observed after 50 passages of some of the cell lines in culture. If elevated β_2-microglobulin expression results from the expression of additional class I molecules by the transformants, this phenotype would be stably expressed because it has been demonstrated previously that the expression of the transferred genes is a stable phenomenon[18].

The ability to measure an elevation in β_2-microglobulin expression by the pooled transformants is related to the fact that in the conditions used for transformation, nearly 100% of the co-transformants express significant and comparable levels

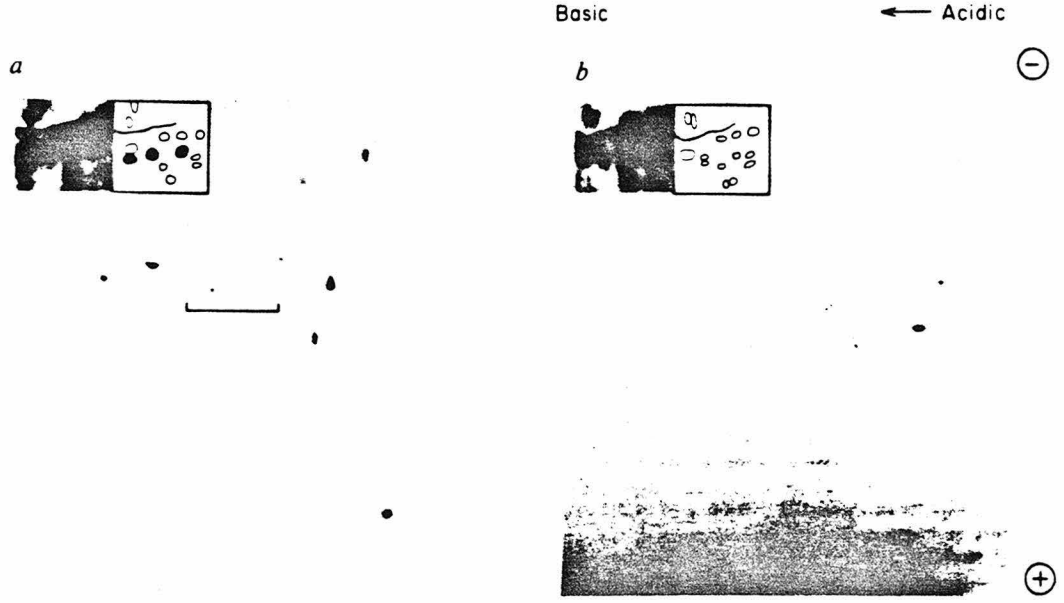


Fig. 7 Fluorographs of two-dimensional gels of the products of the c2.1 gene immunoprecipitated from K8-1 cells (*a*) and tk$^+$ transformants (*b*) using C3H antibodies to K8-1 transformants. The antibodies prepared in C3H mice against c2.1 transformants (K8-1 cells) were used to immunoprecipitate radiolabelled antigens for two-dimensional gel electrophoresis as previously described[18]. The insets contain longer exposures and diagrammatic representations of the 45,000-molecular weight region of each gel. Black spots represent the specific immunoprecipitated molecules from 8-1 cells.

of the foreign class I molecules[36]. It is important to stress that the twofold increase in the magnitude of the radioimmunoassay signal does not necessarily correspond with a twofold increase in the amount of β_2-microglobulin expressed (see ref. 36). These results suggest that the L cells possess the capacity to express additional class I molecules. It has recently been shown that the expression of foreign class I molecules does not occur at the expense of the endogenous H-2^k products of the transformants[36]. Furthermore, we have observed significant levels of expression of as many as three different products by the triple K^d, D^d and L^d cloned transformant line, which demonstrates even greater than a twofold increase in the β_2-microglobulin radioimmunoassay signal.

At least 10 of the transformants which exhibited increased levels of β_2-microglobulin expression failed to react with any of the known serological reagents to class I molecules (Fig. 5*a*,*b*), suggesting that these cells produced novel products associated with β_2-microglobulin. Unfortunately, the heterologous antibodies to β_2-microglobulin proved relatively ineffective in precipitating class I molecules for two-dimensional gel analysis. Thus, we decided to prepare antisera directly against several of these gene products in order to demonstrate formally the presence of novel class I genes. Mouse L cells transformed with the H-2L^d gene have been used to immunize C3H mice, the strain from which the L cell was derived, to generate alloantisera specific for the L^d molecule[20]. Similarly, C3H mice were immunized with L-cell transformants (K8-1) obtained by gene transfer with cosmid clone 2.1, the first gene on cosmid cluster 1 (see Fig. 5*a*). By radioimmunoassay, the antibodies specific for K8-1 cells reacted weakly with BALB/c spleen cells, whereas significantly lower reactivity was observed against BALB/c liver or thymus cells, and C3H spleen, liver or thymus cells (Fig. 6). Accordingly, these antibodies constitute an alloantiserum that was then used to screen spleen cells from a panel of different congeneic and recombinant congeneic strains of mice so that the location of gene c2.1 in the major histocompatibility complex could be assigned by preliminary mapping. These serological analyses suggest that gene c2.1 maps to the right of the complement region. More extensive mapping is required to pinpoint the location of this gene by serological analysis. Such analysis may be complicated if only certain strains of mice express the 2.1 gene products in a manner analogous to that of Qa or TL antigens. However, the data are consistent with the mapping of this gene by restriction enzyme polymorphisms, in that class I gene c2.1 is the first gene in the cosmid cluster containing gene 27.1, which had previously been mapped to the Tla region[15] (Fig. 5*a*).

The antiserum to the K8-1 cells immunoprecipitated molecules from K8-1 cells but not from tk^- transformants or BALB/c spleen cells (Fig. 7). The products of the c2.1 gene appear to be ~45,000 molecular weight and to consist of a number of spots of the appropriate *p*I in a sequential array somewhat characteristic of class I molecules analysed on two-dimensional gels (see Fig. 2). β_2-Microglobulin could not be readily detected in 8-1 immunoprecipitates. This may be due to a low affinity between β_2-microglobulin and the 2.1 gene products, as has been described for certain H-2^d molecules[37]. The origin of the additional products precipitated is unknown. The fact that these products cannot be isolated from spleen cells may be due to the fact that this antigen is expressed at low levels on lymphocytes. However, the ability to generate antibodies to the putative novel gene products expressed by the K8-1 transformants supports the existence of novel genes. Moreover, these genetic, serological and chemical data suggest that gene c2.1 encodes a class I molecule which may, because of its location within the major histocompatibility complex, be a differentiation antigen and not a ubiquitously expressed transplantation antigen.

Mapping of six serologically defined genes

The serologically defined class I genes have been localized within the major histocompatibility complex through the serological analysis of the highly polymorphic congeneic mouse strains. Thus the identification of these class I genes by DNA-mediated gene transfer allows us to map precisely the corresponding cosmid clusters into the major histocompatibility complex. It is interesting that only five class I genes shown in Fig. 5 map to the H-2 region (clusters 2, 11 and 13). Three class I genes encode defined transplantation antigens whereas the remaining two apparently do not encode detectable class I gene products. In contrast, 17 class I genes are contained in cosmid clusters which map to the Tla region. Four class I genes encode serologically defined differentiation antigens and eight encode novel gene products. One of the eight novel genes (c2.1) may encode a differentiation antigen based on the pattern of tissue distribution determined with the alloantibodies. Therefore, the Tla region contains far more class I genes than the H-2 region.

Cosmid clusters 3, 7, 8, 9, 10 and 12 do not express serologically defined gene products and, accordingly, cannot be mapped within the major histocompatibility complex (Fig. 5*b*). These six cosmid clusters contain 15 class I genes, six of which appear to express novel gene products. These clusters are now being mapped through the analysis of restriction enzyme site polymorphisms.

Conclusion

We have identified at least one H-2 K^d, D^d, L^d, possibly a Qa-2,3, and at least two TL genes from the BALB/c mouse. These genes have been mapped into 6 cosmid gene clusters containing 14 class I genes (Fig. 6), of which 9 lie in the Tla region and 5 in the H-2 region. As many as 10 putative novel class I gene products, detected on the basis of their association with β_2-microglobulin, may be encoded by genes dispersed between the mapped and unmapped cosmid gene clusters. In one case, antibodies to a novel class I antigen apparently detect the corresponding antigen expressed at low levels on BALB/c spleen cells.

We thank Drs Shen and Boyse for the gift of the monoclonal antibodies to TL, Drs Hansen and Sachs for the antibodies to the class I antigens, Lorraine Flaherty for the monoclonal antibodies to Qa-2 and Dr Michael Hunkapiller for rabbit antiserum to rat β_2-microglobulin. Dr Michael Steinmetz and Mr Henry Sun are acknowledged for supplying clone DNA, Ms Marie Krempin for assisting with tissue culture, Dr Jeffrey Frelinger for supplying the congeneic mice, Dr Anders Örn for technical assistance, Drs Iwona Stroynowski, Ellen Kraig, Mitch Kronenberg, Ellen Rothenberg, Jane Parnes, Jim Forman and Jeffrey Frelinger for their helpful discussions, and Bernita Larsh for the preparation of the manuscript. This research was supported by NIH grants GM 06965, CA 22662, CA 22199 and CA 25911.

Received 15 July, accepted 20 September 1982.

1. Gorer, P. A. *J. path. Bact* **47**, 231–252 (1938).
2. Gorer, P. A., Lyman, S. & Snell, G. D. *Proc. R. Soc.* **B135**, 499–505 (1948).
3. Zinkernagel, R. M. & Doherty, P. C. *Adv. Immun.* **27**, 51–177 (1980).
4. Shearer, G. M. & Schmitt-Verhurst, A. M. *Adv. Immun.* **25**, 55–91 (1977).
5. Flaherty, L. in *The Role of the Major Histocompatibility Complex in Immunobiology* (ed Dorf, M. E.) 33–58 (Garland STPM Press, New York, 1980).
6. Boyse, E. A., Old, L. J. & Luell, S. *Nature* **261**, 779 (1964).
7. Stanton, T. H. & Boyse, E. A. *Immunogenetics* **3**, 525–531 (1976).
8. Flaherty, L. *Immunogenetics* **3**, 533–539 (1976).
9. Lopez de Castro, J. A. *et al. Biochemistry* **18**, 5704–5711 (1979).
10. Coligan, J. E., Kindt, T. J., Vehara, H., Martinko, J. & Nathenson, S. G. *Nature* **291**, 35–39 (1981).
11. Steinmetz, M. *et al. Cell* **24**, 125–134 (1981).
12. Sood, A. K., Pereira, D. & Weissman, S. M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 616–620 (1981).
13. Ploegh, H. L., Orr, H. T. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6081–6085 (1980).
14. Kvist, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 2772–2776 (1981).
15. Steinmetz, M. S. *et al. Cell* **25**, 683–692 (1981).
16. Malissen, M., Malissen, B. & Jordan, B. R. *Proc. natn. Acad. Sci. U.S.A.* **79**, 893–897 (1982).
17. Evans, G. A., Margulies, D. H., Camerini-Otero, R. D., Ozato, K. & Seidman, J. G. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1994–1998 (1982).
18. Goodenow, R. S. *et al. Science* **215**, 677–679 (1982).

19. Woodward, J. G. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 3613–3617 (1982).
20. Woodward, J. G. *et al. UCLA Symp. molec. cell. Biol.* **24** (1982).
21. Orn, A. *et al. Nature* **297**, 415–417 (1982).
22. Moore, K. W., Taylor-Sher, B., Eakle, K. A., Sun, H. E. & Hood, L. *Science* **215**, 679–682 (1982).
23. Mellor, A. L. *et al. Nature* **298**, 529–533 (1982).
24. Steinmetz, M. S., Winoto, A., Minard, K. & Hood, L. *Cell* **28**, 489–498 (1982).
25. Krakauer, T., Hansen, T. H., Camerini-Otero, R. D. & Sachs, D. H. *J. Immun.* **124**, 2149–2156 (1980).
26. Ivanyi, D. & Démant, P. *Immunogenetics* **8**, 539–550 (1979).
27. Ivanyi, D. & Démant, P. *Immunogenetics* **12**, 397–408 (1981).
28. Démant, P. & Ivanyi, D. *Nature* **290**, 146–149 (1981).
29. Démant, P. & Roos, M. H. *Immunogenetics* **15**, 461–466 (1982).
30. Ivanyi, D. & Démant, P. *Immunogenetics* **15**, 467–476 (1982).
31. Ivanyi, D., Cherry, M. & Démant, P. *Immunogenetics* **15**, 477–454 (1982).
32. Hansen, T. H. *et al. J. Immun* **126**, 1713–1716 (1981).
33. Shen, F.-W., Chorney, M. & Boyse, E. A. *Immunogenetics* (in the press).
34. Vitetta, E. S., Poulik, N. D., Klein, J. & Uhr, J. W. *J. exp. Med.* **144**, 179–192 (1976).
35. Krange, M. S., Orr, H. T. & Strominger, J. L. *Cell* **18**, 979–991 (1979).
36. Goodenow, R. S. *et al. Nature* (submitted).
37. Schmidt, W., Festenstein, H., Ward, P. J. & Sanderson, A. R. *Immunogenetics* **13**, 483–491 (1981).
38. Szybalska, E. H. & Szybalski, W. *Proc. natn. Acad. Sci. U.S.A.* **48**, 2026–2034 (1962).
39. Ozato, K., Mayer, N. M. & Sachs, D. H. *Transplantation* (in the press).
40. Ozato, K., Hansen, T. H. & Sachs, D. H. *J. Immun.* **125**, 2473–2477 (1980).
41. Forman, J., Trial, J., Tonkonogy, S. & Flaherty, L. *J. exp. Med.* **155**, 749–767 (1982).
42. Ozato, K. & Sachs, D. H. *J. Immun.* **126**, 317–321 (1981).

Appendix D

Homologies Between Class I Protein Sequences

The tables in this appendix contain the results of calculations of protein sequence homology between different Class I molecules. Because of the alternative splicing possibilities, only shared sequences were compared. For example, the third external domain of the H-2D^d molecule contains 95 amino acids, but only the 92 amino acids corresponding to the 92 amino acids in the third domain of the H-2L^d protein were compared in the H-2D^d /H-2L^d comparison. Homologies are given as percents. Higher than usual homology values are indicated in boldface. The protein sequences compared are those given in Figure 3 of Chapter II.

Whole Protein (339 residues compared)

	H-2L^d	H-2D^b	H-2K^d	H-2K^b	Q8(27.1)
H-2D^d	85	84	80	86	80
H-2L^d		**95**	82	84	80
H-2D^b			81	82	80
H-2K^d				83	79
H-2K^b					79

First External Domain (∝1)

(90 residues compared)

	H-2L^d	H-2D^b	H-2K^d	H-2K^b	Q8(27.1)	Q10
H-2D^d	80	80	78	**90**	78	77
H-2L^d		**93**	78	84	77	79
H-2D^b			78	83	77	80
H-2K^d				82	82	78
H-2K^b					82	80
Q8(27.1)						79

Second External Domain (∝2)

(92 residues compared)

	H-2L^d	H-2D^b	H-2K^d	H-2K^b	Q8(27.1)	Q10
H-2D^d	91	86	84	85	80	80
H-2L^d		89	86	82	80	85
H-2D^b			83	78	79	82
H-2K^d				80	78	85
H-2K^b					82	80
Q8(27.1)						82

Third External Domain (α3)

(92 residues compared)

	H-2L^d	H-2D^b	H-2K^d	H-2K^b	Q8(27.1)	Q10
H-2D^d	90	90	87	90	88	90
H-2L^d		**100**	90	90	92	91
H-2D^b			90	90	92	91
H-2K^d				87	88	87
H-2K^b					87	86
Q8(27.1)						90

Transmembrane Segment (Exon 5)

(39 residues compared)

	H-2L^d	H-2D^b	H-2K^d	H-2K^b	Q8(27.1)
H-2D^d	69	69	67	64	64
H-2L^d		**100**	67	72	69
H-2D^b			67	72	69
H-2K^d				**82**	64
H-2K^b					74

First Cytoplasmic Segment (Exon 6)

(11 residues compared)

	H-2L^d	H-2D^b	H-2K^d	H-2K^b	Q8(27.1)
H-2D^d	100	100	82	100	73
H-2L^d		100	82	100	73
H-2D^b			82	100	73
H-2K^d				82	62
H-2K^b					77

Second Cytoplasmic Segment (Exon 7)

(13 residues compared)

	H-2L^d	H-2D^b	H-2K^d	H-2K^b	Q8 (27.1)
H-2D^d	85	85	77	85	77
H-2L^d		100	62	69	61
H-2D^b			62	69	61
H-2K^d				92	64
H-2K^b					64