

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

BY DOUGLAS A. FISHER, STEPHEN W. HUNT III, AND LEROY HOOD

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

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

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

In the past four years, our understanding of class I genes has increased dramatically due to the isolation of complementary DNA (cDNA) clones from both human and mouse class I genes (9–11). Steinmetz et al. (12, 13) used a class

This work was supported by grants from the National Institutes of Health, Bethesda, MD. S. Hunt is a Leukemia Society postdoctoral fellow.

¹Abbreviations used in this paper: cDNA, complementary DNA; dNTP, deoxyribonucleoside triphosphate; HPLC, high-pressure liquid chromatography; mAb, monoclonal antibody; MHC, major histocompatibility complex; mRNA, messenger RNA; TL, thymic leukemia.

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

In this report, we present the complete genomic nucleotide sequence and analysis of a BALB/c gene (*Tla^c*) encoding a TL antigen. The polypeptide chain is encoded in six exons homologous to the first six exons of other class I genes. Southern blot analyses using *Tla*-specific probes subcloned from this gene have allowed us to reorganize the *Tla*-region gene clusters described earlier (13), and have reduced the number of class I genes detected in the BALB/c mouse to 33: 18 in the *Tla* region, 10 in the *Qa-2,3* region, and 5 in the *H-2* region.

Materials and Methods

Materials. Restriction endonucleases and T4 DNA ligase were obtained from New England Biolabs, Beverly, MA. The large fragment of *E. coli* DNA polymerase was from Bethesda Research Laboratories, Gaithersburg, MD. Deoxynucleoside triphosphates (dNTP) were obtained from Sigma Chemical Co., St. Louis, MO, dideoxynucleoside triphosphates and the 15-nucleotide sequencing primer from Collaborative Research (Waltham, MA), and α -[³²P]dNTP from Amersham Corp. (Arlington Heights, IL). *E. coli* strain JM103, phage M13mp8, and phage M13mp10 were from Bethesda Research Laboratories.

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

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

Results and Discussion

Gene 17.3A Encodes a Serologically Defined TL Antigen. Clone 17.3 was previously shown (15) to transform mouse L cells to the TL⁺ phenotype, as assayed

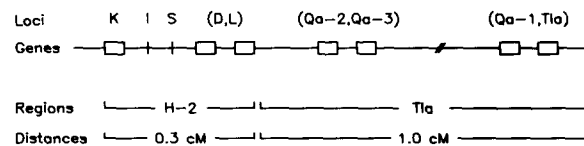


FIGURE 1. Genetic map of the MHC. Genes indicated with open boxes encode class I molecules. The order of loci within brackets is not known. Distances are in centimorgans (cM).

by anti-TL mAb in a radioimmunoassay. A partial restriction map of the eukaryotic insert of clone 17.3 is shown in Fig. 2. Since clone 17.3 contains two class I genes (denoted 17.3A and 17.3B), it was necessary to determine which of the genes was responsible for the TL-transforming activity. Digestion of clone 17.3 with the restriction endonucleases Sac II or Bam HI, which cut within 17.3B, did not affect the TL-transforming activity, whereas digestion with Eco RI, which cleaves only within 17.3A, abolished the TL-transforming activity (R. Goodenow, California Institute of Technology, unpublished results). Also shown in Fig. 2 is pTLA.1, a pBR322 subclone derived from 17.3 that contains only gene 17.3A. Subclone pTLA.1 contains the TL-transforming activity of clone 17.3, and this activity is abolished by digestion with Eco RI. Since Eco RI digests the insert DNA only once, within gene 17.3A, we conclude that gene 17.3A encodes a serologically defined TL antigen.

The 17.3A Polypeptide Chain is Encoded by Six Exons Homologous to the First Six Exons of Transplantation Antigen Genes. The nucleotide sequence of gene 17.3A was determined using the sequencing strategy shown in Fig. 2. The DNA sequence is shown in Fig. 3. The first five exons of gene 17.3A correspond both

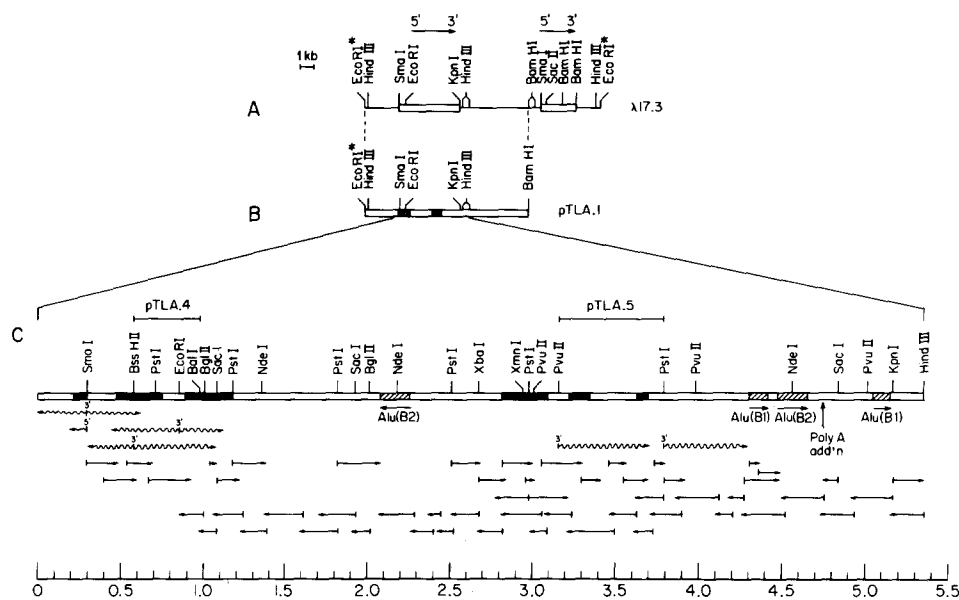


FIGURE 2. Restriction map and sequencing strategy for gene 17.3A. (A) Partial restriction map of the insert of Charon 4A clone 17.3A. Open boxes indicate regions hybridizing with class I cDNA clones (11), and the arrows indicate the transcriptional orientation of the two class I genes, 17.3A and 17.3B.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

in sequence and in the position of intron-exon borders with the first five exons of other class I genes. These exons encode a hydrophobic leader peptide, three external protein domains of about 90 amino acids each, and a hydrophobic transmembrane segment. A cytoplasmic protein domain is encoded entirely by exon 6, unlike transplantation antigens, where the cytoplasmic domain is usually encoded by three exons.

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

Exons 2, 3, and 4 encode three protein regions of ~90 amino acids each, and are denoted $\alpha 1$, $\alpha 2$, and $\alpha 3$, respectively (33). These three regions comprise all of the TL molecule that is exposed on the cell surface. The positions of cysteine residues are homologous to those of other class I genes: both the $\alpha 2$ and $\alpha 3$ regions contain cysteine residues ~60 amino acids apart, suggesting that each contains an intrachain disulfide linkage, while the $\alpha 1$ region contains no cysteines. The TL molecule encoded by gene 17.3A contains two potential carbohydrate attachment sites (Asn-X-Ser_{Thr}), both in the $\alpha 1$ domain, at amino acids 86 and 90. It is not known from biochemical data how many carbohydrate moieties are present on TL molecules from BALB/c (*Tla^c*) mice, but there is only one moiety on TL molecules encoded by the *Tla^a* allele (34).

Exon 5 encodes the transmembrane segment of the TL polypeptide. The translated amino acid sequence contains 38 amino acids, including a span of 20 uncharged and generally hydrophobic amino acids (15 hydrophobic, 3 glycine, and 2 serine residues) that presumably span the lipid bilayer and terminate in a cytoplasmic region with two positively charged amino acids.

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

The low sequence homology between transplantation antigens and gene 17.3A

Tla GENE SEQUENCE AND ORGANIZATION

1 AAGTCCCGGAAGTGGTCGTCAGGGTCTGAGCCTGAAGGCGGGGATAGGTTGGGCTGTACAGAAGTATAGTTTCACITTTTGCCTCTCTG 90
91 TACCATTGTGCAGATTGTCTTCAAAGGCCCTGCACTCTGCCGAGAGCAGGTAAGTGAAGATGCAAACTACTCTGAGTTGCAAGITTA 180

-26 EXON 1 M R M G T Q V P S T L L -14
181 CAAAGTCCAAGACTTTATACAACCTTCGGATCTCCCTAAC ATG AGG ATG GGG ACA CAA GTG CCT AGC ACC CTC CTC 255

-14 I L L A A S Q V Q T Q T C P -1
256 ATC CTC CTG GCT GCC TCA CAA GTC CAG ACC CAG ACC TGC CCG G GTGAGTGTGGATGAGGAGGAAACAGCCTTT 330
331 TAGGGGAGGGGGGAGGTTCCATCTCCACCAGACCCCTCCACCCTTCTCCACTCAAGGCCCTAGACCTGAGTCCCTTCTACTTCCCTGGCC 420

1 EXON 2 G S H S L R Y F 8
421 TGCACACCCCTGCCGGTCTGAGGAGGAGTCTCGAAGTCTCACACGCGCCGCCCCAG GC TCA CAC TCG CTG AGG TAC TTC 500

9 Y T A L S R P A I S E P W Y I A V G Y L O D 30
501 TAC ACC GCC TTG TCC CGA CCT GCA ATC AGC GAA CCA TGG TAC ATC GCT GTG GGC TAC CTG GAT GAC 566

31 T Q F A R F D S A G E T G T Y K L S A P W V 52
567 ACT CAG TTC GCG CGC TTC GAC AGC GCA GGG GAG ACT GGG ACA TAT AAG CTA AGT GCG CCA TGG GTG 532

53 E Q E G P E Y W A R E T E I V T S N A Q F F 74
633 GAG CAA GAG GGG CCC GAG TAT TGG GCG AGA GAG ACA GAG ATC GTC ACA AGC AAT GCA CAG TTT TTC 698

75 R E N L Q T M L D Y Y N L S Q N 90
699 CGC GAG AAT CTG CAG ACT ATG CTG GAC TAC AAC CTG AGT CAA AAT G GTGAGGAACCCCAAAATCTAAGA 771
772 GTCTGTCCCACTCTCTCCCTCTTATGCAAGCCACCTAATCATGGGGTCAAAAAGGGACCGGTTTCCCTTTCAGTTTGGAGAATCTCTA 861

91 EXON 3 G S H T I Q V M Y G C E V E 104
862 GGGTGGGCGGGTCTGACTAGGAGAGGGGGCCAG GC TCT CAC ACC ATC CAG GTG ATG TAT GGC TGT GAG GTG GAG 935

105 F F G S L F R A Y E Q H G Y D G Q D Y I A L 126
936 TTC TTC GGG AGC CTC TTC CGC GCT TAT GAG CAG CAT GGC TAT GAT GGC CAA GAT TAC ATC GCC CTG 1001

127 N E D L K T W T A A D M A A E I T R S K W E 148
1002 AAT GAA GAT CTG AAA ACG TGG ACA GCA GCA GAC ATG GCA GCA GAG ATC ACC AGA AGC AAG TGG GAG 1067

149 Q A G Y T E L R R T Y L E G P C K Q S L L R 170
1068 CAG GCT GGT TAT ACA GAG CTC CGT AGG ACA TAC TTG GAG GGC CCA TGC AAG GAT TCC CTG CTC AGA 1133

171 Y L E N R K K T Q E C T 182
1134 TAC CTG GAG AAC AGA AAA AAG ACA CAG GAG TGC ACA G GTGGGTGGCTGCAGGCAACTCCATCTCCCTTTGCCCTCTGT 1210

1211 ACTGAGTTCAGTCCAGAGAGAGCAGAACTACTAGCTGGGAAAACCTCCATCTCTTTTTCTTTTTTGTGTTGTTGTTGCTG 1300
1301 TTTGGTGGTTTTTTTGTCTTTTAAATTTTAAAAGGTGACAACAGAGAAATAGCATATGGAAATACACAAAAGCAAAGGCTATACCCAGCT 1390
1391 CAATAACTTATAAACANTTAAANNCCCACTTTGGACCASTAGAGGATATGPAATTTTACATATTTTCCATTAATAATACCTCATGCTATCC 1480
1481 AACATCGACCCAGGGGTGAGGCGAGGAGAGAGACAGAGACAGAGACTCACASACACAGAGAGACAGAGGCGAGAGGCGAGAAAGAGATGG 1570
1571 ASAGACAGAAAGAGAGAAAAATCAGTGGCAGAATTAAGAAGTACATCAATTTATGCCCTGCCATTAATCCATTCGGTCCCTTACCTACAT 1660
1661 TTGAGTCCCTGGCCATCTCCAAGTATAGCCCTTCCCTGATTTAATTTATGTCATGTTGCAAGTAACAACAGGTCACATTTCTAAACAAGTAG 1750
1751 TACGTTTTTTTGTGCTTTGTTCTTTGTTGTTGTTTGGCTAAATACCTCTATGAGCAAAAAGACTGCAGTTTTACTTCCCTGTTGTT 1840
1841 GTTCTGAGATGGGTGGACTTTGTTGAATTTGTATATTCAAAAGTTCATCAATTTTCTAGATAAATGTTGATTACATTTGTTAGAGCTC 1930
1931 AACAGGATTCACCTTTATATATATACCCCATATTCGTTAGTCACTTAAGTGAATGCTGGGTTGCTTTACTGCAATAGCTAAGATCTCACA 2020
2021 CTCATATCAATTAAGAGAGTAGAGAGTGGCACAGTCTTTCTCTCTCTTTCTTTTTTTAAGATTTATTTATTTATTATATA 2110
2111 CATTAAAGTAACTGTAGCTTCTTCAGACACACCAGNAGGTGTCAGTCTCATTACAGATGGTTGTGAGCCACCATATGGTTGCTGGAAATTTG 2200
2201 AACTCAGGACCTTCGAAGAGCAATCAGTGTCTTACCCATTGAGCCATCTCACCAGCCCTGGTAGTTCTTTGTTATTTCGGTATCTTACCA 2290
2291 GCAATACTGGGTGTSTCTTCAATTTGACTCAAGGTTCAAGCTCTTCAATGTTCTCTTCTTCAATATGGGAGGGGGAGAGCACTCCTGTGT 2380
2381 TTCTGACTCTCTCAGAGACTGCCCTTCCAAGGAGCACCTTTGTCTCTGGACAGTTTCAGATTTCTTTCAGGATGAAAAGGGAGAGAGAA 2470
2471 GGTTCTGAAAGANNNAAAAACATTTCCCATGAGTCTGCASTGCTTTTGTAGCTATGGCTTCTCTCAGGCCCTTGTCTCTGCCACACCCA 2560
2561 ACCTCTGTGACATTTGACTTCAAGTGTGCTGAATCTCTGAGCGCTACCCAGACAGAAAGCAGACGCTCCCTTTTTGGCTCTGGGACACATA 2650
2651 GCTCTCTGGGCAAGTTCTAGAATTTTCAAGCATTTTATGTAACCTACTGGGAAAATTTGACCCCTTGTTCAGTCAATTTCTTCCC 2740

183 EXON 4 D P P K 186
2741 AGGTAGCCTCACAGACACTTCTGGAGTCAACAGGAAAAATATATGATGCCCAACTTTCCAACCTTCTCTCTCAG AT CCT CCA AAA 2826

187 T H V T H H A R P E G D V T L R C W A L G F 208
2827 ACA CAT GTG ACC CAT CAC GCC AGA CCT GAA GGT GAT GTC ACC CTG AGG TGC TGG GCC CTG GGC TTC 2892

209 Y P A H I T L T W Q L N G E E L I Q D T E L 230
2893 TAC CCT GCT CAC ATC ACC CTG ACC TGG CAG TTG AAT GGG GAG GAA TTG ATT CAG GAC ACG GAG CTT 2956

FIGURE 3.

231	V E T R P A G D G T F Q K W A A V V V P S G	252
2359	GTG GAG ACC AGG CCT GCA GGG BAT GGA ACC TTC CAG AAG TGG GCA GCT GTG GTG GTG CCT TCT GGG	3024
253	E E Q K Y T C H V Y H E G L P E P L T L R W	274
3025	GAA GAG CAG AAA TAC ACA TGT CAT GTG TAC CAT GAG GGG CTG CCT GAG CCT CTC ACC CTG AGA TGG	3090
275		275
3091	G GTAAGGAGGGGTGTGGGTGCAGAGCTGGGGTCAAGGAAAGCTGGAGCCTTCTGAAGACCTCAGCTGGTCAGATTGAGAGTTGCGATC	3179
275		285
3180	ATGTTGCTCACTTTCCCATCATGTCTTACCCTTCCCTTCCCAAG AG CCT CCT CAG TCC AGT ATG CCC AAC AGG ACC	3255
288	T V R A L L G A M I I L G F M S G S V M M W	307
3256	ACT GTT CGT GCT CTC CTT GGA GCT ATG ATC ATC TTA GGT TTT ATG AGC GGA AGT GTT ATG ATG TGG	3321
308	M R K N N	312
3322	ATG AGA AAG AAC AAT G GTATGGAAGAGTCTGTGGCTGGGAGCCTTATGATTGTAAACCAATACACATGCACCTCATGCAAAAAGA	3405
3406	TTGCCTATCTCTGGGTCTGTTTGCAGACACTCACCTTACAATAAAGACAGATTTCATAATGGTAATGGGGTTCTGCACAGGATTCTAAA	3495
3496	TAACCTCTCAGAGGTAAGGGCAAGGCTCATGCTAAACACTACAGTCTCCAGACTGGCCTTGTTCATCCCTGCATATAATGTCCATG	3585
313		323
3586	GCACAGAACTTCCCTGGTCTCTCATATCTTTTTTTTTTTTTTAAACAG GT GGA AAC GGA GAC GAT AAC ACT GCT GCA TAT	3664
324	Q N E R E H L S L T S G X	335
3665	CAG AAT GAG AAG GAA CAC TTG TCC ACC TCG GGC TGA <u>ATCTGAGGCACAGG</u> GTGGAAGCTGGGATGAAGGATC	3741
3742	TTCCCTCAGCCCCACCCTGCTCTGCAATGTCTTGTACTTCTACTGAGGCAAGGACAGCTTCTGCTCTGCTGGAGACTTGA	3831
3832	<u>AAAG</u> GTGACACCTGGGATTGGGGCTGAGGCAAACTGGAAGACTCGGTCACTTGGCCCTTTCATAATTAACCTTTGAATGAGAGGAAA	3921
3922	CCTTTGAGATGTTTTACATTGACTGCCCTGACTCTTTTTCTTCCCTTGGGACAGCTGCCACATGGGACTGAGTCACAAGTTCTGACTT	4011
4012	CTGAACAGTGAATCACTTTGGGACAGTGTCTGTGTTTTATGCTTTTGACTGTGCATACATGGBTAATCTGGAGAGCACAGCCTGCC	4101
4102	CCTGCACACCAAGACTCAGACCCCTACATTAGCCATCATCCCTTGACACBCCAACCTTCTCGGCTCTGGGACATCTTCATCC	4191
4192	TGTCAGCTCCACGCTGCTGCTGGACCTGTGCTCCTCACACGCCACAGTGTATCTCCCTCATTTCCACACTGAACATGAAATCTGAATGTG	4281
4282	GATGTAGGAGTGGTGGTGGGCCATTAATCCTAGCAGCTGGGAGGCAGAGGCAGGCAGATTCTGAGTTCGAGGCCAGCCTGGTCTACAA	4371
4372	AGTGAGTTCACAGGACGCCAGGCTACACAGAGAAATCTATTCTTGGAAAAAACAACAAGGGAAAAAAGAAAAAAGAAAAAAGAATC	4461
4462	TGAATGTGGAGGCTGGAGPBGAYGGCTCAGTGGTGTAAACGCTACTGCTCTTCTGAGGACCTGATTTCAATCCCCAGCAACACATGGTGG	4551
4552	CTCACAAACATATGCAATAGCATCCAATGCCCTTCTCGGGTGTGTGGAAGACACCTACAGTGTATTCCACTACATGAAATAAATAATC	4641
4642	TTAGAAAAAGTATGAGTGTGAACCTCTTGACCTGAGCAGTTCACCTTGTGGGTAATTAAGGATTGGGAATATCTTAATATTGTGGAGGA	4731
4732	<u>ATAAAT</u> GTAGGACCTTCCAGAATCTGTGTTCTTGTGCTGAGTATGTCAGGTGGGGCAGGAGTCTCTGGCATCTGCTGTGAAATGGG	4821
4822	TGTGCCAGGTGGAGCTCAGTCCATGTGCAGTCTTGTGCTGGTCTGGTCAATTCCTCAGCTATACCTCTGTGGTACTTGTCACTTAATAGA	4911
4912	GTCACTGGATGACAGACTCTGATGATGGTAACACGAAATGCCCTTACCTTCTCTTATGACCGSACTATGAGCCCCGTAPACTGTACTAGG	5001
5002	CCTGTTGTGCAGCTGAAAACTGACATTACATTAAAAAAATTTCTGGAGAGTAGTGTGAACACCTTTAATCCCAACATTCAGGAGACAG	5091
5092	GCAGGAAAAACCACTGATTTCTAGGCATCCTGGACACTCAGGAGTGTGACACAGAGAAACCTGTCTCCAAATGGGTACCTGGGAGCAGGGTA	5181
5182	TTCCCTGTGCTCAGCCTGACCATGCAGATTGTAGGCAGAAAGTAGATTTGGGTTTTGGATTAGAAGAACGTTGCATGCAGGAAGCAAGGC	5271
5272	TTAAAGAGCCAAGCTAGTGGAGGCTGGGAGGCAGAGCTGAGGGCAGGTTGACCTTTGTGGGCCAAGGTAATGAAAGCTT	5350

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

in the cytoplasmic region may be a clue to the different behavior of TL molecules and transplantation antigens in the cell membrane. In a process known as antigenic modulation, TL antigens are rapidly lost from the cell surface after incubation with anti-TL antibody (even monovalent antibody), while transplantation antigens are not (35-37). The presence of a very different cytoplasmic domain suggests that TL antigens may interact with the cytoskeleton or other

cytoplasmic components differently than H-2 antigens, but further studies, such as exon-shuffling experiments, will be needed to localize the structural component responsible for the antigenic modulation effect.

Comparison of the Nucleotide and Amino Acid Sequence of 17.3A with K^d, D^d, L^d, and a Qa-2,3 Gene. The DNA sequence of gene 17.3A was compared with sequences from the following class I genes: K^d (17), L^d (16), D^d (20), the Qa2,3 gene Q6 (27.1) (12), K^b (18), and the Q10 gene, encoding a secreted class I molecule from the Qa-2,3 region (19, 38). None of these genes is significantly more closely related to gene 17.3A than to the other, and all are more closely related to each other than any one is to gene 17.3A. With an overall homology to the K^d gene of only 74% at the DNA level, gene 17.3A is the most divergent murine class I gene sequenced to date. Table I shows the DNA and amino acid homologies of gene 17.3A with four BALB/c class I genes: K^d from the K end of the H-2 locus, L^d and D^d from the D end of H-2, and gene Q6, from the Qa-2,3 region.

All four genes are approximately equally homologous to gene 17.3A, but each exon is not equally conserved (Table I). The fourth exon is the most highly conserved exon, probably due to selective pressure on the sequence of the $\alpha 3$ region encoded by exon 4, since the $\alpha 3$ region binds noncovalently with β_2 -microglobulin (39). Exons 1 and 5, encoding the hydrophobic leader and transmembrane peptides, are only moderately conserved, since the sequence constraints on these two regions probably require only hydrophobic properties. The cytoplasmic exon is the least conserved and, as mentioned above, this may reflect different interactions with cytoskeletal elements. The degree of homology of exons 2 and 3 with the genes in Table I is intermediate between that seen for exon 4 and exon 5. In H-2 molecules, the $\alpha 1$ and $\alpha 2$ regions are the site of recognition by cytotoxic T lymphocytes (40), but since the function of the $\alpha 1$ and $\alpha 2$ regions in TL antigens is unknown, it is difficult to estimate the nature of the selective pressure operating on these regions.

The homologies noted above are consistent with earlier peptide mapping analyses. Peptide mapping studies showed that out of 13–19 tryptic peptides resolved on a high-pressure liquid chromatography (HPLC) cation exchange column, three peptides coeluted between the H-2D^d molecule and a Tla^a gene

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

Region of 17.3A	K ^d		L ^d		D ^d		27.1 (Qa-2,3)	
	DNA	Protein	DNA	Protein	DNA	Protein	DNA	Protein
	%							
Exon 1 (leader)	64	52	53	43	64	54	56	29
Exon 2 ($\alpha 1$)	71	56	70	56	72	58	70	56
Exon 3 ($\alpha 2$)	74	62	75	65	74	65	75	63
Exon 4 ($\alpha 3$)	91	83	92	85	92	89	91	86
Exon 5 (TM)	64	35	64	38	65	41	62	35
Cytoplasmic	42	35	42	30	43	30	45	31
Total	74	61	74	61	75	64	74	60

product, for an estimated peptide homology of 20%. In identical experiments, different H-2 antigens were found to have ~40% peptide map homology, and allelic forms of TL were found to share 70–80% of their peptides (41, 42). The translated amino acid sequences of the D^d and 17.3A genes share three predicted tryptic peptides. Two of the peptides come from the $\alpha 3$ region and one is the N-terminal peptide. The actual percent peptide homology between the D^d and TL antigens is somewhat lower than the previous estimates because the HPLC system used to separate the tryptic peptides could only resolve 16–19 peptides, while 30–43 peptides are predicted from the translated amino acid sequence.

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

The Third Intron of Gene 17.3A Contains Large DNA Deletions and/or Insertions. The third intron is related in an interesting manner to the third intron of the K^d gene (Fig. 4A). At a point 95 bp from the 5' end of this intron, gene 17.3A contains 1,093 bp of sequence not found in the K^d gene. Whether this is an insertion in gene 17.3A or a deletion in the K^d gene is unknown, but the 1,093-bp segment contains a 65-bp thymidine-rich (76%) sequence at its 5' end, and an *Alu*-like repeat element near its 3' end. The segment is flanked by inexact 10-bp inverted repeats, which are also present in the K^d gene (Fig. 4B). The

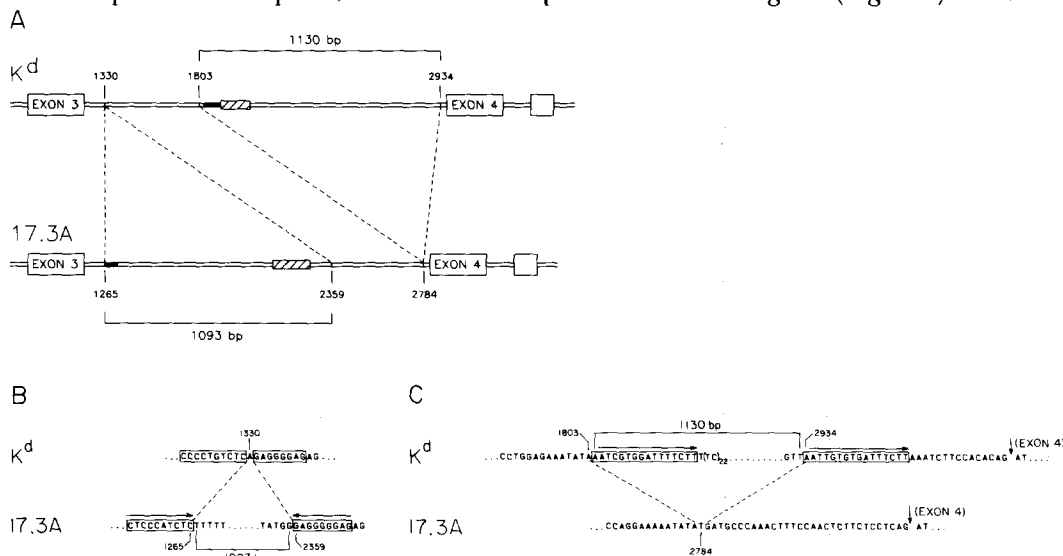


FIGURE 4. Noncoding sequences of gene 17.3A. (A) Diagram showing the relationship between the third intron of the K^d gene and the 17.3A gene. Open boxes indicate exons, hatched boxes indicate *Alu* repetitive elements, and the solid regions in the introns (near 1,803 in the K^d gene and 1,265 in 17.3A) indicate simple sequences. (B) DNA sequence in the K^d and 17.3A genes at the borders of the 1,093-bp sequence found in 17.3A. (C) DNA sequences bordering the 1,130 bp sequence in the K^d gene.

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

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

To test whether the fourth exon of gene 17.3A is more highly conserved than would be expected from natural selection on the protein sequence, we counted the number of silent site mutations (base substitutions that do not alter the protein sequence) between the L^d and 17.3A genes. Silent site mutations accumulate with time at a constant rate, and are believed to be independent of selection at the protein level, so by using their frequency, one can estimate the time elapsed since two related genes shared a common ancestor (43). There are mutations in 43% of the silent sites in exons 2 and 3 of gene 17.3A but only in 11% of the silent sites of exon 4 when compared with the L^d gene (Fig. 5B). This disparity (43% vs. 11%) is very large, and corresponds to a statistically significant (>3 SD) difference in divergence times estimated for these different regions by the calculation of Kimura (43). This means that the fourth exon of gene 17.3A shared a common ancestor with the L^d gene much more recently than the second and third exons. This last finding, coupled with the observation

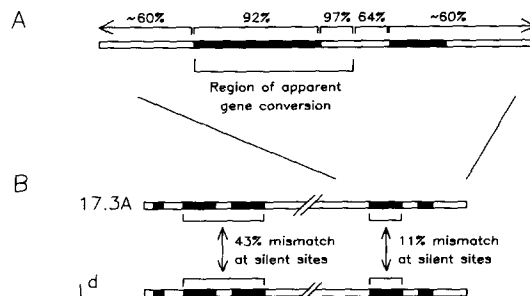


FIGURE 5. Region of apparent gene conversion in 17.3A. (A) Exons 4 and 5 are shown as solid boxes, and introns are shown as open boxes. Numbers at the top indicate the percent nucleotide homology with the L^d gene in the bracketed areas. (B) Diagram showing the silent-site comparisons made between 17.3A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

that the first 71 bp of the fourth intron are highly conserved, leads us to the hypothesis that a region including all of exon 4 and part of the fourth intron was transferred to the 17.3A gene as a gene conversion event from one of the *H-2* genes. A gene conversion of this size (~350 bp) would serve to decrease the sequence polymorphism of the fourth exon, in contrast to the role envisioned for very small (<50 bp) gene conversions in exons 2 and 3 of mutant *H-2* genes which, it has been suggested (18, 44, 45), would increase the polymorphism of these regions.

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

Gene conversions of variable size occur between class I genes, but it is unclear how many of these events are biologically significant. The small gene conversions mentioned above may contribute to the polymorphism of transplantation antigens. However, it is difficult to imagine why natural selection would not be sufficient to maintain the protein sequence of the $\alpha 3$ domain. Gene conversion may be a result rather than a cause of the high DNA sequence homology of exon 4. Since gene conversions presumably require an initial base pairing between homologous sequences, conversions may occur more frequently in exon 4 because of the high sequence homology of this exon with other class I genes.

Hybridization With Tla-specific Probes Detects a Small Subset of Class I Genes That Are Polymorphic Between Tla Alleles. To study the number and diversity of genes related to 17.3A, as well as to create DNA probes suitable for examining RNA species expressed in TL^+ cells, two subclones, pTLA.4 and pTLA.5, were constructed from the coding regions of gene 17.3A that exhibited minimal homology with other class I genes (Fig. 2). The 5' probe, pTLA.4 (bases 577–892), contains parts of the second and third exons, as well as the entire second intron. This probe is ~70% homologous with the K^d or L^d genes in the coding sequences. The 3' probe, pTLA.5 (bases 3,155–3,795), contains the transmembrane exon and is only ~50% homologous to the corresponding regions in the K^d and L^d genes. Both probes hybridize to BALB/c thymus poly(A)⁺ RNA, but not to BALB/c liver or spleen RNA. The probes also do not hybridize to thymus, liver, or spleen RNA from C57BL/6, a TL^- strain (data not shown).

Subclones pTLA.4 and pTLA.5 were hybridized to genomic DNA from four strains of mice representing the *Tla* alleles Tla^a , Tla^b , Tla^c , and Tla^d . As can be seen in Fig. 6, there is size polymorphism in the hybridizing restriction fragments. This is surprising since others have had relative difficulty finding polymorphic restriction fragments in the *Tla* region compared with the *H-2* region (14). Four *Tla*-specific bands are found in the Southern analysis of Tla^c mouse DNA. These data are summarized in Table II. One band corresponds to gene 17.3A, the

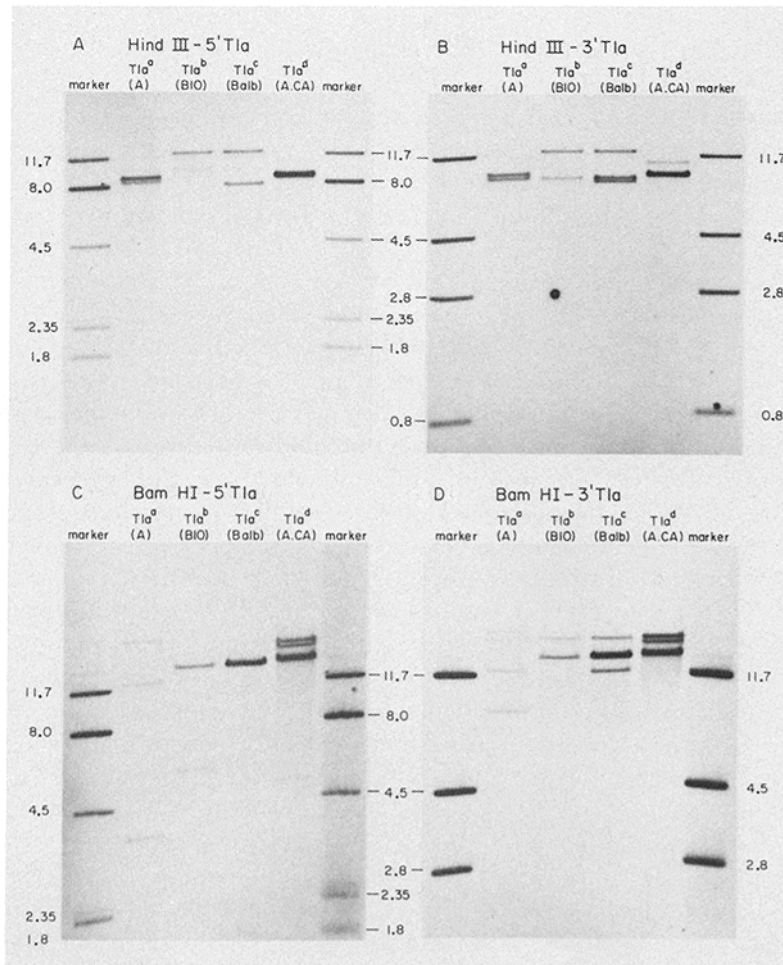


FIGURE 6. Genomic southern blots with *Tla*-specific probes on DNA from four mouse strains. $10 \mu\text{g}$ (5×10^{-18} moles) of genomic DNA from the indicated strains was digested to completion with Hind III or Bam HI, electrophoresed through a 0.8% agarose gel in $1 \times \text{TAE}$ (0.04 M Tris acetate pH 7.5, 0.002 M EDTA) blotted onto nitrocellulose and hybridized 12–18 h with 5×10^5 cpm/ml of the indicated nick-translated ($1-4 \times 10^8$ cpm/ μg) probes. The final wash was in $1 \times \text{SSC}$ (0.15 M NaCl, 0.015 M sodium citrate, pH 7.0) at 68°C . Hybridization markers are restriction enzyme–digestion M13 subclones diluted to 10^{-17} moles per fragment. (A) Hind III–digested DNA hybridized with pTLA.4. (B) Hind III–digested DNA hybridized with pTLA.5. (C) Bam HI–digested DNA hybridized with pTLA.4. (D) Bam HI digests hybridized with pTLA.5. The bands in the *Tla*^a lane in C and D appear less intense than the other lanes because only 3–5 μg of DNA was loaded in those lanes. This inaccuracy was corrected in A and B. All bands in the *Tla*^c lane do not hybridize with equal intensity. Those bands corresponding to gene 24.8 and the 24.8-like gene (genes T1 and T11 in Fig. 7; see also Table II) are less intense in B and D. The hybridizing region of 24.8 is only 85% homologous to pTLA.5, and would be expected to hybridize more weakly than the 100% homologous 17.3A gene (D. Fisher, unpublished results).

gene from which the probes were subcloned, and another corresponds to gene 24.8, the other *Tla* gene identified by transformation into L cells (15). However, gene 24.8 by itself could not encode a TL antigen because DNA sequence

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

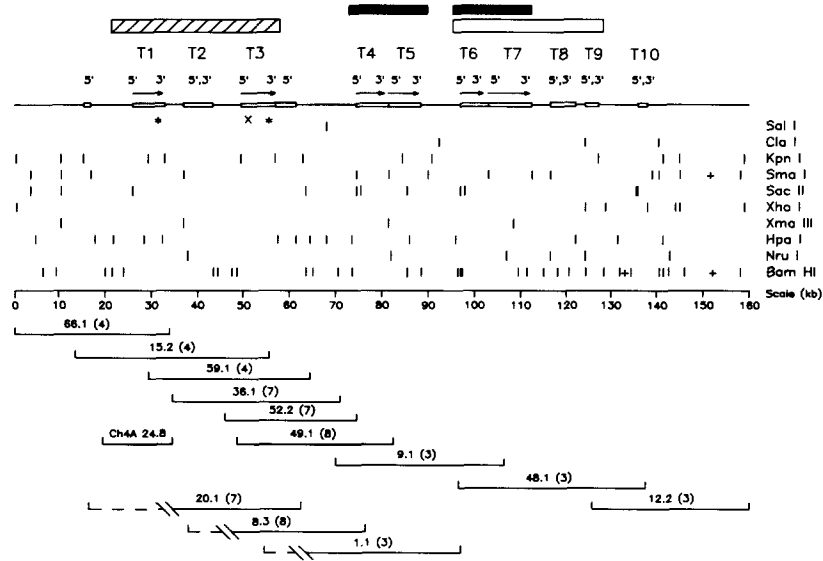
Gene name	Gene in Fig. 7	pTLA.4	pTLA.5	Size of hybridizing fragment		
				Hind III	Bam HI	Kpn I
17.3A-like	T3	+	+	13	15.5	7.8
17.3A	T13	+	+	8	15.5	7.8
24.8	T1	-	+	8.3	~20	3.7
24.8-like	T11	-	+	8.3	12	3.7

analysis indicates that this gene is, in fact, a pseudogene, due to the presence of numerous in-frame stop codons (D. Fisher, unpublished results). Gene 17.3A hybridizes with both the 5' and 3' *Tla* probes, while gene 24.8 hybridizes only to the 3' probe. The other two hybridizing genes are of two types: one is like 17.3A because it hybridizes to both pTLA.4 and pTLA.5, while the other is 24.8-like and hybridizes only to pTLA.5.

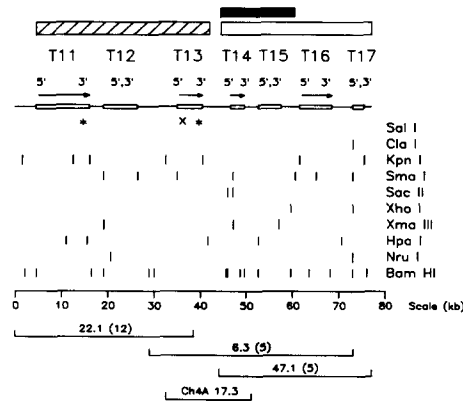
New Linkage of Tla Clusters. When the two *Tla* probes were hybridized to a panel of cosmid clones containing all 36 class I genes of the gene clusters described by Steinmetz et al. (13), seven genes, rather than the expected four, were detected. This observation prompted us to examine these cluster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

Elimination of the artifactual constructs allows several new overlaps between the gene clusters described by Steinmetz et al. (13) (Fig. 7). Clusters 3, 7, 8 and 4 can now be joined into a single cluster 160 kb in length, which we denote cluster A. Cluster A contains the three artifactual constructs. A second cluster of 77 kb, denoted cluster B, is formed from clusters 12 and 5. This linkage was previously overlooked because the 10-kb overlapping region contains only three mapped restriction sites. A third *Tla* cluster, cluster 10, a single cosmid clone containing an incomplete class I gene, is unchanged from the earlier study. This new linkage is consistent with genome blots using both *Tla*-specific probes, because four genes (T1, T3, T11, and T13) hybridize to pTLA.5, and two genes (T3 and T13) hybridize to pTLA.4 (Table II; Figs. 6 and 7). The sizes of the restriction fragments hybridizing to these probes, as determined in genomic blots (Table II), are consistent with the sizes predicted from the restriction maps of

A



B



C

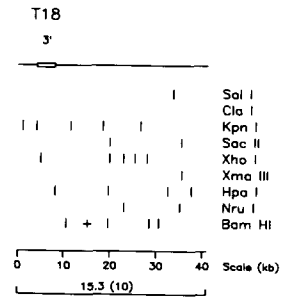


FIGURE 7. Class I *Tla* genes of the BALB/c mouse. (A) Restriction map and organization of gene cluster A. Regions by hybridizing with class I cDNA probes (11) are indicated as thin open boxes. 5' (pH-2III)- or 3' (pH-2IIa)- hybridizing regions are indicated above the boxes, and the transcriptional orientation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

the corresponding cosmid clones. Similar cluster linkage results have recently been independently obtained by others (49). The new gene linkage reduces the number of class I genes found in the BALB/c mouse to 33: 18 in the *Tla* region, 10 in the *Qa-2,3* region, and five in the *H-2* region.

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

Tla^b Mice Contain Fewer Tla Genes Than Tla^c Mice, and Gene 17.3A Is One of Those Absent in Tla^b. Recently, Weiss et al. (48) have isolated cosmid clones containing class I genes from the C57BL mouse (*Tla^b*), and have linked the *Tla*-region class I genes into one gene cluster. 26 class I genes were found in C57BL as compared to 33 in BALB/c. Most of the additional BALB/c genes were found in the *Tla* region. The *Tla^b* gene cluster can be aligned with cluster A, and all the genes in cluster A are found in the *Tla^b* cluster (Fig. 8). However, no genes corresponding to cluster B or cluster 10 are found among the C57BL cosmid clones. Genomic Southern blots (Fig. 6) using the *Tla*-specific probes confirm that at least two genes (genes T11 and T13 in Fig. 7) in BALB/c are not found in *Tla^b*, and both of these genes are located in cluster B. Although these two genes are the only ones confirmed to be absent in *Tla^b*, it seems probable that none of the genes in cluster B or cluster 10 are present in C57BL mice.

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



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

gene and its homolog in *Tla^c* are expressed in leukemic cells, while 17.3A (absent in *Tla^b*) is expressed in thymocytes.

Summary

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

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

Received for publication 4 April 1985.

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

References

1. Zinkernagel, R. M., and P. C. Doherty. 1980. MHC-restricted cytotoxic T cells: studies on the role of polymorphic major transplantation antigens determining T-cell restriction-specificity, function, and responsiveness. *Adv. Immunol.* 27:51.
2. Old, L. J., E. A. Boyse, and E. Stockert. 1963. Antigenic properties of experimental leukemias. I. Serological studies in vitro with spontaneous and radiation-induced leukemias. *J. Natl. Cancer Inst.* 31:977.
3. Flaherty, L. 1980. The Tla region antigens. In *Role of the Major Histocompatibility Complex in Immunology*. M. E. Dorf, editor. Garland Press, New York. 33-57.
4. Cook, R. G., and N. F. Landolfi. 1983. Expression of the thymus leukemia antigen by activated peripheral T lymphocytes. *J. Exp. Med.* 158:1012.
5. Scheid, M. P., G. Goldstein, and E. A. Boyse. 1978. The generation and regulation of lymphocyte populations: evidence from differentiative induction systems in vitro. *J. Exp. Med.* 147:1727.
6. Lepault, F., and I. Weissman. 1981. An in vivo assay for thymus-homing bone marrow cells. *Nature (Lond.)* 293:151.
7. Shen, F., M. Chorney, and E. A. Boyse. 1982. Further polymorphism of the Tla locus defined by monoclonal antibodies. *Immunogenetics.* 15:573.
8. Michaelson, J., E. A. Boyse, M. Chorney, L. Flaherty, E. Fleissner, U. Hämmerling, C. Reinisch, R. Rosenson, and F.-W. Shen. 1983. The biochemical genetics of the Qa-Tla region. *Transplantation Proc.* 15:2033.
9. Ploegh, H. L., H. T. Orr, and J. L. Strominger. 1980. Molecular cloning of a human histocompatibility antigen cDNA fragment. *Proc. Natl. Acad. Sci. USA.* 77:6081.
10. Sood, A. K., D. Pereira, and S. M. Weissman. 1981. Isolation and partial nucleotide sequence of a cDNA clone for human histocompatibility antigen HLA-B by use of an oligodeoxynucleotide primer. *Proc. Natl. Acad. Sci. USA.* 78:616.

11. Steinmetz, M., J. G. Frelinger, D. Fisher, T. Hunkapiller, D. Pereira, S. M. Weissman, H. Uehara, S. Nathenson, and L. Hood. 1981a. Three cDNA clones encoding mouse transplantation antigens: Homology to immunoglobulin genes. *Cell*. 24:125.
12. Steinmetz, M., K. W. Moore, J. G. Frelinger, B. T. Sher, F. Shen, E. A. Boyse, and L. Hood. 1981b. A pseudogene homologous to mouse transplantation antigens: Transplantation antigens are encoded by eight exons that correlate with protein domains. *Cell*. 25:683.
13. Steinmetz, M., A. Winoto, K. Minard, and L. Hood. 1982. Clusters of genes encoding mouse transplantation antigens. *Cell* 28:489.
14. Winoto, A., M. Steinmetz, and L. Hood. 1983. Genetic mapping in the major histocompatibility complex by restriction enzyme polymorphisms: Most mouse class I genes map to the Tla complex. *Proc. Natl. Acad. Sci. USA*. 80:3425.
15. Goodenow, R. S., M. McMillan, M. Nicolson, B. T. Sher, K. Eakle, N. Davidson, and L. Hood. 1982. Identification of the class I genes of the major histocompatibility complex by DNA mediated gene transfer. *Nature (Lond.)*. 300:231.
16. Moore, K. W., B. T. Sher, Y. H. Sun, K. A. Eakle, and L. Hood. 1982. DNA sequence of a gene encoding a BALB/c mouse L^d transplantation antigen. *Science (Wash. DC)*. 215:679.
17. Kvist, S., L. Roberts, and B. Dobberstein. 1983. Mouse histocompatibility genes: structure and organization of a K^d gene. *EMBO (Eur. Mol. Biol. Organ.) J.* 2:245.
18. Weiss, E., L. Golden, R. Zakut, A. Mellor, K. Fahrner, S. Kvist, and R. A. Flavell. 1983. The DNA sequence of the H-2K^b gene: evidence for gene conversion as a mechanism for the generation of polymorphism in histocompatibility antigens. *EMBO (Eur. Mol. Biol. Organ.) J.* 2:453.
19. Mellor, A. L., E. H. Weiss, M. Kress, G. Jay, and R. A. Flavell. 1984. A nonpolymorphic class I gene in the major histocompatibility complex. *Cell*. 36:13.
20. Sher, B. T., R. Nairn, J. Colligan, and L. Hood. 1985. DNA sequence of the mouse D^d gene. *Proc. Natl. Acad. Sci. USA*. 82:1175.
21. Davis, R. W., D. Botstein, and J. R. Roth. 1980. *Advanced Bacterial Genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
22. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular cloning, a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
23. Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe. 1980. Cloning in single stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol.* 143:161.
24. Maxam, A. M. and W. Gilbert. 1980. Sequencing end labeled DNA with base specific chemical cleavages. *Methods Enzymol.* 65:499.
25. Frischauf, A. M., H. Garoff, and H. Lehrach. 1980. A subcloning strategy for DNA sequence analysis. *Nucleic Acids Res.* 8:5541.
26. Hong, G. F. 1982. A systematic DNA sequencing strategy. *J. Mol. Biol.* 158:539.
27. Barnes, W. B., M. Bevan, and P. H. Son. 1983. Kilo-sequencing: Creation of an ordered nest of asymmetric deletions across a target sequence carried on phage M13. *Methods Enzymol.* 101:98.
28. Davis, M. M., K. Calame, P. W. Early, D. L. Livant, R. Joho, I. L. Weissman, and L. Hood. 1980. An immunoglobulin heavy chain gene is formed by at least two recombinational events. *Nature (Lond.)*. 283:733.
29. Krayev, A. S., D. A. Kramerov, K. G. Skryabin, A. P. Ryskov, A. A. Bayev, and G. P. Georgiev. 1980. The nucleotide sequence of the ubiquitous repetitive DNA sequence B1 complementary to the most abundant class of mouse fold-back RNA. *Nucleic Acids Res.* 8:1201.
30. Krayev, A. S., T. V. Markusheva, D. A. Kramerov, A.P. Ryskov, K. B. Skryabin, A.

- A. Bayev, and G. P. Georgiev. 1982. Ubiquitous transposon-like repeats B1 and B2 of the mouse genome: B2 sequencing. *Nucleic Acids Res.* 10:7461.
31. Kozak, M. 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res.* 12:857.
 32. Breathnach, R., and P. Chambon. 1981. Organization and expression of eukaryotic split genes coding for proteins. *Ann. Rev. Biochem.* 50:349.
 33. Coligan, J. E., T. J. Kindt, H. Uehara, J. Martinko, and S. G. Nathenson. 1981. Primary structure of a murine transplantation antigen. *Nature (Lond.)* 291:35.
 34. Rothenberg, E., and D. Triglia. 1981. Structure and expression of glycoproteins controlled by the Qa-1^a allele. *Immunogenetics.* 14:455.
 35. Boyse, E. A., L. J. Old, and S. Luell. 1963. Antigenic properties of experimental leukemias. II. Immunological studies in vivo with C57BL/6 radiation-induced leukemias. *J. Natl. Cancer Inst.* 31:987.
 36. Boyse, E. A., E. Stockert, and L. J. Old. 1967. Modification of the antigenic structure of the cell membrane by thymus leukemia (TL) antibody. *Proc. Natl. Acad. Sci. USA.* 58:954.
 37. Stackpole, C. W., J. B. Jacobson, and M. P. Lardis. 1974. Antigenic modulation in vitro. I. Fate of thymus leukemia (TL) antigen-antibody complexes following modulation of TL antigenicity from the surfaces of mouse leukemia cells and thymocytes. *J. Exp. Med.* 140:939.
 38. Kress, M., D. Cosman, G. Khoury, and G. Jay. 1983. Secretion of a transplantation related antigen. *Cell* 34:189.
 39. Yokoyama, K., and S. G. Nathenson. 1983. Intramolecular organization of class I H-2 MHC antigens; localization of the alloantigenic determinants and the β_2 -m binding site to different regions of the H-2 K^b glycoprotein. *J. Immunol.* 130:1419.
 40. Stroynowski, I., A. Orn, R. S. Goodenow, M. McMillan, J. Forman, P. R. Braxton, J. Frelinger, and L. Hood. 1984. Cytotoxic T lymphocytes recognize determinants on the BALB/c-H-2L^d molecule controlled by α 1 and α 2 but not α 3 external domains. *Immunogenetics.* 20:141.
 41. Yokoyama, K., E. Stockert, L. J. Old, and S. G. Nathenson. 1981. Structural comparisons of TL antigens derived from normal and leukemia cells of TL⁺ and TL⁻ strains and relationship to genetically linked H-2 major histocompatibility complex products. *Proc. Natl. Acad. Sci. USA.* 78:7078.
 42. McIntyre, K. R., U. Hämmerling, J. W. Uhr, and E. S. Vitetta. 1982. Structural analysis of thymus leukemia (TL) antigens in the mouse. *J. Immunol.* 128:1712.
 43. Kimura, M. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. *Proc. Natl. Acad. Sci. USA.* 78:454.
 44. Weiss, E. H., A. Mellor, L. Golden, K. Fahrner, E. Simpson, J. Hurst, and R. A. Flavell. 1983b. The structure of a mutant H-2 gene suggests that the generation of polymorphism in H-2 genes may occur by gene conversion like events. *Nature (Lond.)* 301:671.
 45. Pease, L. R., D. H. Schulze, G. M. Pfaffenbach, and S. G. Nathenson. 1983. Spontaneous H-2 mutants provide evidence that a copy mechanism analogous to gene conversion generates polymorphism in the major histocompatibility complex. *Proc. Natl. Acad. Sci. USA.* 80:242.
 46. Sodoyer, R., M. Damotte, T. L. Delovitch, J. Trucy, B. R. Jordan, and T. Strachan. 1984. Complete nucleotide sequence of a gene encoding a functional human class I histocompatibility antigen (HLA-CW3). *EMBO (Eur. Mol. Biol. Organ.) J.* 3:879.
 47. Strachan, T., R. Sodoyer, M. Damotte, and B. R. Jordan. 1984. Complete nucleotide sequence of a functional class I HLA gene, HLA-A3: implications for the evolution of HLA genes. *EMBO (Eur. Mol. Biol. Organ.) J.* 3:887.

48. Weiss, E. H., L. Golden, K. Fahrner, A. L. Mellor, J. J. Devlin, H. Bullman, H. Tiddens, H. Bud, and R. A. Flavell. 1984. Organization and evolution of the class I gene family in the major histocompatibility complex of the C57BL/10 mouse. *Nature (Lond.)* 310:650.
49. Rogers, J. H. 1985. Family organization of mouse H-2 class I genes. *Immunogenetics*. 21:343.