MOLECULAR EVIDENCE THAT THE *H-2D* AND *H-2L* GENES AROSE BY DUPLICATION

Differences between the Evolution of the Class I Genes in Mice and Humans

By RONALD J. RUBOCKI,* DAVID R. LEE,‡ WEN-RONG LIE,* NANCY B. MYERS,* AND TED. H. HANSEN*

*From the *Department of Genetics, Washington University School of Medicine, St. Louis, Missouri 63110; and the ‡Department of Molecular Microbiology and Immunology, University of Missouri-Columbia School of Medicine, Columbia, Missouri 65212*

The class I genes of the murine MHC encode glycoproteins that are noncovalently associated with β_2-microglobulin (β_2m)[1]. These class I genes are located in four regions of the MHC, the K, D, Q, and TL regions. The classical transplantation antigens of the K and D regions function by presenting endogenous peptides derived from processed self or foreign proteins to the TCR (1), while the function of the Q and TL molecules is not known. In addition, the K and D region molecules display a high level of polymorphism, whereas few alleles have been demonstrated for the Q and TL class I genes (2). The high level of polymorphism in the transplantation antigens is due to amino acid substitutions concentrated in or near the peptide binding groove, as shown by recent crystallization data (3). Analyses of the K and D region class I genes from wild-type and mutant genes suggest that gene conversion-like events may play a major role in the generation of diversity (4). This high level of diversity is thought to increase the immune repertoire of individual members as well as the species as a whole.

Although the number of K region class I genes and products has been found to be constant in different haplotypes, the number of D region class I genes and products varies in different haplotypes (5, 6). Haplotype disparity involving the D region genes has been demonstrated in both inbred laboratory and wild-derived mouse strains using serological and peptide mapping analyses (7–11). This disparity has been investigated at the molecular level in four haplotypes, b, k, p, and d (6, 12–14). Of these haplotypes, all have a single D region gene, except the d haplotype, which contains five D region class I genes (12). However, it is not clear whether there are only these two D region organizations. Evidence suggesting that this region represents a dynamic state of expansion and contraction is provided by the dm1 and dm2 mutant mouse strains. Molecular studies of these mutants revealed deletions of several

This work was supported by National Institutes of Health Public Health Service grants AI-23410, AI-19687, and AI-27123.

Address correspondence to Ted. H. Hansen, Department of Genetics, Washington University School of Medicine, Box 8232, 4566 Scott Avenue, St. Louis, MO 63110.

[1] *Abbreviations used in this paper:* β_2m, β_2-microglobulin; FIGE, field inversion gel electrophoresis.

D region class I genes, most likely resulting from unequal crossover events (15, 16). Another unique property of the D region is that there appears to be a family of highly homologous class I gene products, with the L^d molecule being the prototypic structure (10). This suggests an apparent low level of polymorphism, which may have implications on the selective advantage of having an L^d-like molecule on the cell surface.

To resolve issues regarding the evolution of D region class I genes and their relationship to other class I-encoding regions of the mouse, as well as man, we have characterized the D region of the B10.AKM mouse strain. Several previously published findings suggested that an analysis of the D^q region would yield important new information. Sequential immunoprecipitation analyses defined two molecules encoded by the D^q region, designated D^q and L^q (7, 8); and recent molecular studies identified the D^q and L^q genes that, when expressed in L cells, functioned as typical transplantation antigens (17). Nucleotide sequence comparisons of the 5′ halves of the D^q and L^q genes indicated that both were strikingly similar to the L^d gene sequence (17). However, the restriction map flanking the D^q gene appeared to align with the D^d locus, whereas the L^q gene appears to align with the L^d gene. Thus, we speculated that the D^q region would provide key information regarding a putative duplication event that resulted in multiple class I genes within the *D* region. To verify this assumption, we considered it critical to isolate all of the class I genes encoded within the D^q region and compare these genes to previously characterized class I genes.

In this report, we describe the isolation of cosmid clones containing the D^q, $D2^q$, $D3^q$, $D4^q$, and/or $Q1^q$ genes. Furthermore, we report the completion of the entire sequence of the D^q and L^q genes, as well as the 5′ half of the $D4^q$ gene. Our findings provide several insights into how the class I genes in the D region of the mouse evolved. Furthermore, significant differences are identified between the evolution of the class I gene family in mice and humans.

Materials and Methods

DNA Isolation. DNA was isolated from B10.AKM (K^k, I^k, D^q) mouse liver using a modified procedure of Chang et al. (18). Briefly, mouse livers were homogenized in 5% citric acid and then overlayed onto a 5% citric acid/30% sucrose cushion. The isolated nuclei were lysed with an SDS-proteinase K mixture and then sequentially extracted with phenol/chloroform and ether. The DNA was concentrated with solid phase sucrose for a few hours and then dialyzed extensively against 10 mM Tris-HCl, pH 7.5, 1 mM EDTA.

Generation and Screening of Genomic Libraries. Genomic B10.AKM liver DNA was partially digested with Mbo I under conditions to optimize the yield of fragments in the size range of 35–40 kb. The digested DNA was then size fractionated by electrophoresis on a Bull's eye gel apparatus (Hoefer Scientific Instruments, San Francisco, CA). Cosmid vector pTCF, which contains the selectable aminoglycosyl-3′-phosphotransferase gene, was linearized with Bam HI and dephosphorylated with calf intestine alkaline phosphatase (Boehringer Mannheim Biochemicals, Indianapolis, IN). The 35–40-kb genomic inserts were ligated into linearized pTCF cosmid vector using T4 ligase (International Biotechnologies, Inc., New Haven, CT). The ligated concatamers were packaged into λ phage using Gigapack-plus packaging extracts (Stratagene, San Diego, CA), and the cosmid library was screened by a modification of the protocol described by Hanahan and Meselson (19). Primary, secondary, and tertiary screenings of the cosmid library were performed using a 3′ flanking probe isolated from the region flanking the L^d gene (L^d 3′ Flc [see probe 4 in reference 16]) and a 3′ noncoding region probe

from the K^q gene (pH8D [20]). Additional probes used were a Bgl II-Pst I $\alpha 3$ probe isolated from the L^d gene, a 6.2-kb Hpa I probe from cosmid II 2.20 (12), a 1.4-kb Bam HI probe from cosmid 50.2 (12), and a 2-kb Sal I probe from B10.AKM cosmid 33.3. These probes were labeled by the random hexamer priming technique of Feinburg and Vogelstein (21) using α-[^{32}P]dATP and α-[^{32}P]dCTP (Amersham Corp., Arlington Heights, IL). In addition, five oligonucleotide probes were used to further screen the isolates: $L^d\alpha 3A$ (5'-TTCACCTTTAGATCTGGGGTGATC-3'), complementary to the 5' variable region in the $\alpha 3$ exon of the L^d gene; L^dM (5'-TCCAATGATGGCCATAGC-3'), complementary to the variable region of the transmembrane exon of the L^d gene; $D2^d$M, $D3^d$M, and $D4^d$M, complementary to the transmembrane exon variable region in the $D2^d$, $D3^d$, and $D4^d$ genes, respectively (obtained from S. Hunt, University of North Carolina, Chapel Hill, NC) (22). The oligonucleotide probes were labeled at the 5' end with T4 kinase and γ-[^{32}P]ATP (New England Nuclear, Boston, MA).

Restriction Endonuclease Mapping and Alignment of Cosmid Clones. Recombinant cosmids purified by cesium chloride banding were used for restriction endonuclease site mapping according to the method of Graham et al. (23). Briefly, recombinant cosmid DNA was linearized with restriction endonuclease Cla I, which cuts in the vector. The linearized cosmid clones were partially digested with restriction endonucleases (0.1–0.2 U/μg DNA) and the fragments generated were resolved by field inversion gel electrophoresis (FIGE) on a 1% agarose gel. Electrophoresis was performed for 20 h at 12°C in 0.5× TBE (89 mM Tris-Borate, 2 mM EDTA) at 350 V in the forward direction and 250 V in the reverse, with a switching time of 0.3 s. These conditions were optimal for resolution of fragments up to 50 kb. After blotting the gel onto reinforced nitrocellulose (Micron Separations, Inc., Westborough, MA), the filter was hybridized at 65°C with random hexamer-labeled probes (21) that correspond to the left (1.4-kb Hind III–Hpa I) or right ends (1-kb Cla I–Bgl II) of the linearized vector. The blot was washed twice at 65°C in 0.5× SSC for 30 min and exposed to X-Omat AR film for 1 h with a single intensifying screen. The restriction endonuclease sites were calculated from the size of the partial fragments that hybridized to either the left or right end probes. The resulting restriction maps were then compared with each other, and overlapping cosmids were grouped together. The alignment of these D^q region cosmids with the previously mapped D^d region of BALB/c was performed by comparing the restriction maps for Cla I, Sal I, Nru I, Bam HI, Kpn I, and Hpa I. The location within the overlapping cosmid clones of the six isolated genes was determined by DNA blotting and hybridization with the class I gene probes previously described.

DNA Sequencing. The 5' Bam HI–Xba I fragments (1.4 kb) from the D^q and L^q genes containing sequence from exon 1 through intron 3 were sequenced as previously described (17). The 3' Bam HI fragment from D^q was subcloned into the Bluescript plasmid vector. The 3' half of the L^q gene was subcloned from cosmid clone 33.3 by digestion with the restriction endonuclease Bam HI. Double-stranded DNA containing the 3' half of D^q and L^q were sequenced by dideoxy sequencing (24) using sequence-specific primers, 5' α[^{32}P]dATP (ICN, Irvine, CA; and Amersham Corp.), and Sequenase (U. S. Biochemical, Cleveland, OH). The fragments were sequenced in both directions. Primers for use in the sequencing of the L^q-cosmid construct were provided by Dr. David Chaplin (Washington University, St. Louis, MO).

The 5' half of $D4^q$ was sequenced by subcloning the two Bam HI fragments, comprising exons 1–3, individually into Bam HI–digested M13mp18 or M13mp19. Sequencing was performed using the dideoxy method (24) with [^{35}S]ATP (Amersham Corp.) and Sequenase.

Genomic DNA Blot Analyses. High molecular weight DNA was prepared according to a modified procedure of Chang et al. (18). 10–15 μg of digested DNA was fractionated by electrophoresis in 0.7% agarose gels containing ethidium bromide at 0.5 μg/ml and then transferred to reinforced nitrocellulose. Prehybridization was performed for 2 h in 6× SSC/5× Denhardt's solution, 20 mM sodium phosphate buffer (pH 7.0), 4 mM EDTA, 200 μg of heat-denatured salmon sperm DNA/ml, 0.2% $NaDodSO_4$ at 68°C. Hybridizations were performed overnight at 68°C with probes labeled by random hexamer priming (21). After hybridization, the nitrocellulose was washed for 1 h at 68°C in 0.5× SSC. The blots were then exposed to X-Omat AR film between two intensifying screens. Subsequently, 0.1× SSC was used to wash the blot for an additional 1 h at 68°C, after which the blot was reexposed to film.

Results

Isolation of the D^q, $D2^q$, $D3^q$, $D4^q$, L^q, and $Q1^q$ Genes, and Alignment to BALB/c Class I Genes. In an earlier study, we reported the cloning of the D^q and L^q genes from a cosmid library of B10.AKM genomic DNA. The restriction maps of clones containing these genes showed a striking similarity to the published restriction maps of clones containing the D^d and L^d genes, respectively (17). Based on this observation and genomic DNA blot comparisons of other groups (12), the D^q and D^d regions were predicted to have similar genetic organizations. To test this assumption, low copy probes were derived from D^d region cosmids, and their D region specificity was confirmed using DNA from the deletion mutants dm1 and dm2. The dm1 mutant expresses one D region gene that is a hybrid of the D^d and L^d genes (15), while the dm2 mutant expresses only D^d as a result of a deletion comprising $D2^d$, $D3^d$, $D4^d$, and L^d (16). The location of the D region-specific probes used for this study is shown in Fig. 1. Each of these probes was found to hybridize with DNA from B10.AKM and was used for isolation of cosmid clones from the D^q region. In addition to these genomic probes, oligonucleotide probes specific for the transmembrane segments of the $D2^d$, $D3^d$, or $D4^d$ genes were used. The hybridization patterns of these D^d region-derived probes on B10.AKM cosmid clones are summarized in Table I. Restriction maps for each of these clones were generated by FIGE of partial digestions (23). This approach eliminated the need to subclone smaller fragments. The sizes of the restriction fragments and the number of sites were also confirmed by complete digests of each of the clones. Based on restriction map similarities and/or hybridization with the panel of probes, the D^q-derived clones were aligned with the previously published map of the D^d region (Fig. 1). Progressing in a telomeric direction, cosmid clones 5.1, 3.1, 34.2, and 95 were found to contain the D^q gene. Each of these clones was transfected into a murine L cell line (DAP-3) and expressed a serologically indistinguishable D^q molecule (17, and data not shown). Cosmid clone 91 was aligned to the telomeric end of the D^q gene based on overlapping restriction fragments and its hybridization with probe II2.20. This probe, previously described by Stephan et al. (12), hybridizes with two restriction fragments in BALB/c DNA, one flanking the L^d gene, and the other flanking the D^d gene. Probe II2.20 also detects two fragments in B10.AKM DNA. The fragment in clone 91 was found to be the same size as the D^d-specific fragment from BALB/c. Cosmid clone 6 contains two class I genes that crosshybridize to the $D2^d$M and $D3^d$M oligonucleotide probes. In addition, probe 50.2A, isolated from the BALB/c cosmid clone, 50.2 (containing the $D2^d$ and $D3^d$ genes), also crosshybridized with cosmid clone 6 DNA. Thus, cosmid clone 6 contains $D2^q$ and $D3^q$. In both BALB/c and B10.AKM, the *D3* gene was localized to a 12-kb Kpn I fragment. This map position in the d haplotype differs somewhat from an earlier report (12). Overlapping cosmid clones 17 and 33.3 (previously shown to also contain the L^q gene [17]) both hybridized with the $D4^d$M probe, indicating that they contain the $D4^q$ gene. The remaining cluster of cosmid clones 36, 53, and 38 were all localized to the telomeric side of the L^q gene. This map position is based on their crosshybridization with probe II2.20. This probe, as mentioned above, detects both a *D*- and *L*-specific RFLP, and clones 36, 53, and 38 each contain the *L*-specific fragment. The latter two cosmid clones, 53 and 38, were found to contain a class I gene provisionally designated $Q1^q$.

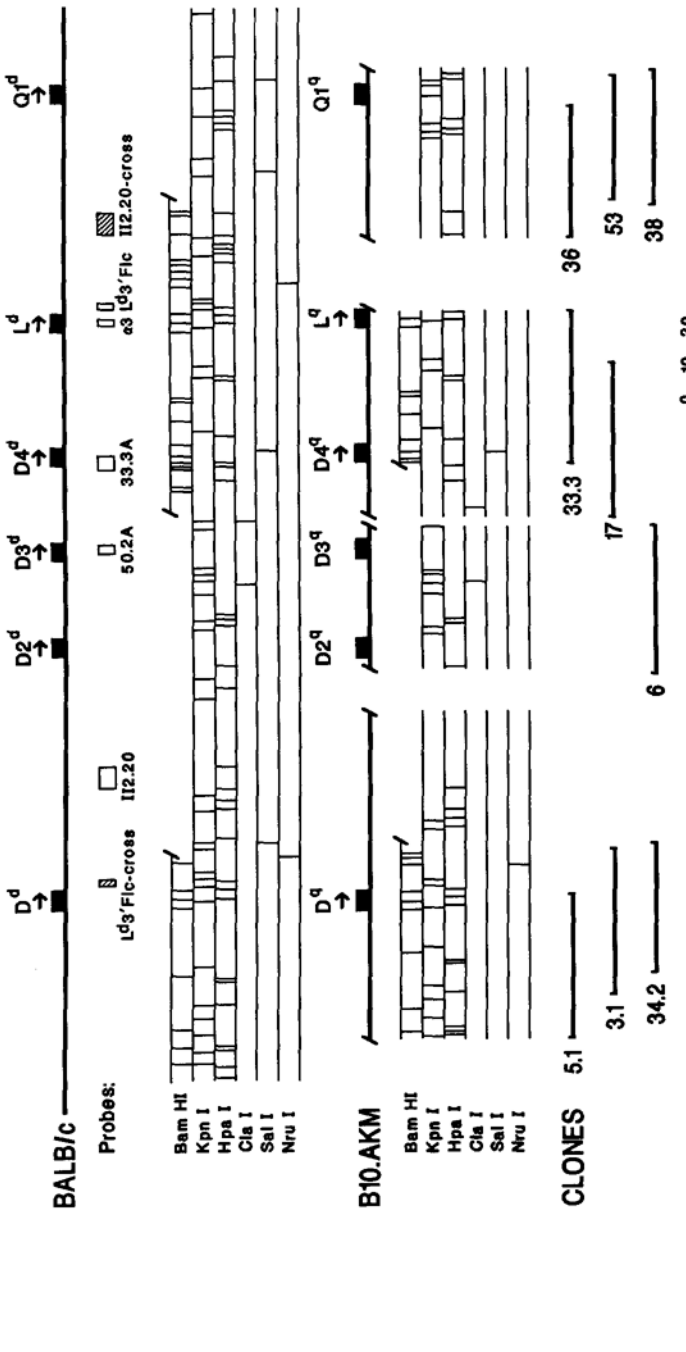

FIGURE 1. Alignment of B10.AKM *D* region clones with BALB/c (12, 15). The $Q1^q$ gene cluster was aligned based on crosshybridization with probe II2.20. The $D2^q$ and $D3^q$ genes were aligned based on restriction endonuclease sites and crosshybridization with $D2^d$ and $D3^d$ oligonucleotide probes. The location of probes is shown by open boxes, while their crosshybridizations are shown by hatched boxes.

TABLE I

Hybridization Patterns of B10.AKM Cosmid Clones

	Probes									
Clones	$L^d\alpha 3A$	L^dM	II2.20	50.2A	D2M	D3M	D4M	$\alpha 3$	33.3A	$L^d3'Flc$
3.1*	+	+	–	–	–	–	–	+	+	+/–‡
5.1*	+	+	–	–	–	–	–	+	+	+
34.2*	+	+	–	–	–	–	–	+	+	+
95	+	+	+	–	–	–	–	+	+	+
91	–	–	+	–	–	–	–	–	–	–
6	–	–	–	+	+	+	–	+	+	–
17	–	–	–	–	–	–	+	+	+	–
33.3*	+	+	–	–	–	–	+	+	+	–‡
36	–	–	+	–	–	–	–	–	–	–
53	–	–	–	–	+	–	–	+	+	–
38	–	–	–	–	+	–	–	+	+	–

* Reported previously.
‡ Truncated gene.

Interestingly, DNA from cosmid clones 53 and 38 also hybridized strongly with the $D2^dM$ probe, suggesting a common origin of the $D2^q$ and $Q1^q$ genes. Collectively, these findings demonstrate that the D^q region has a very similar restriction map and gene organization as the D^d region.

Sequence Comparisons of $D4^q$, D^q, and L^q. The nucleotide sequence of the 5′ half of the D^q and L^q genes was previously published (17), whereas the sequence of the 3′ half is shown in Fig. 3. In addition, the sequence representing the 5′ portion of the $D4^q$ gene, comprising exons 1–3, is shown in Fig. 4. It is readily apparent that D^q and L^q are very homologous to each other and to L^d. The leader sequences of D^q and L^q are identical to L^d, whereas L^q has one nucleotide and D^q has eight nucleotide differences, as compared with the $\alpha 1$ exon of L^d (17). These differences result in identical $\alpha 1$ domains of the L^q and L^d molecules, and five amino acid substitutions between the D^q and L^d molecules (Fig. 5). The $\alpha 2$ exons of D^q and L^q have the most differences, in comparison with L^d, of any stretch within these genes. The D^q gene has 10 nucleotide substitutions resulting in eight amino acid differ-

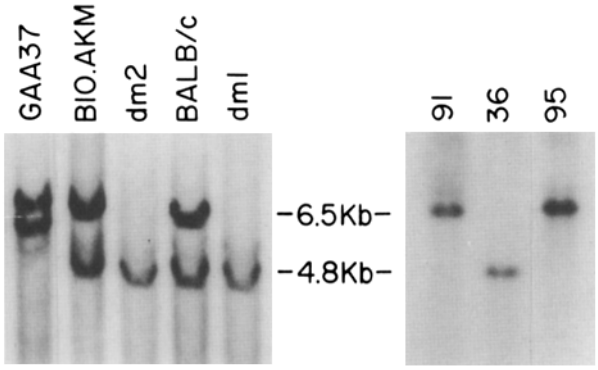

FIGURE 2. DNA blot analysis of B10.AKM cosmid clones and genomic DNA. Cosmid clones 36, 91, and 95 in addition to B10.AKM, dm2, BALB/c, and dm1 genomic liver DNA were digested with Bam HI, blotted, then hybridized to probe II2.20. Bands at 6.5 and 4.8 kb correspond to fragments mapped ~25 kb 3′ of D^d and 3′ of L^d, respectively.

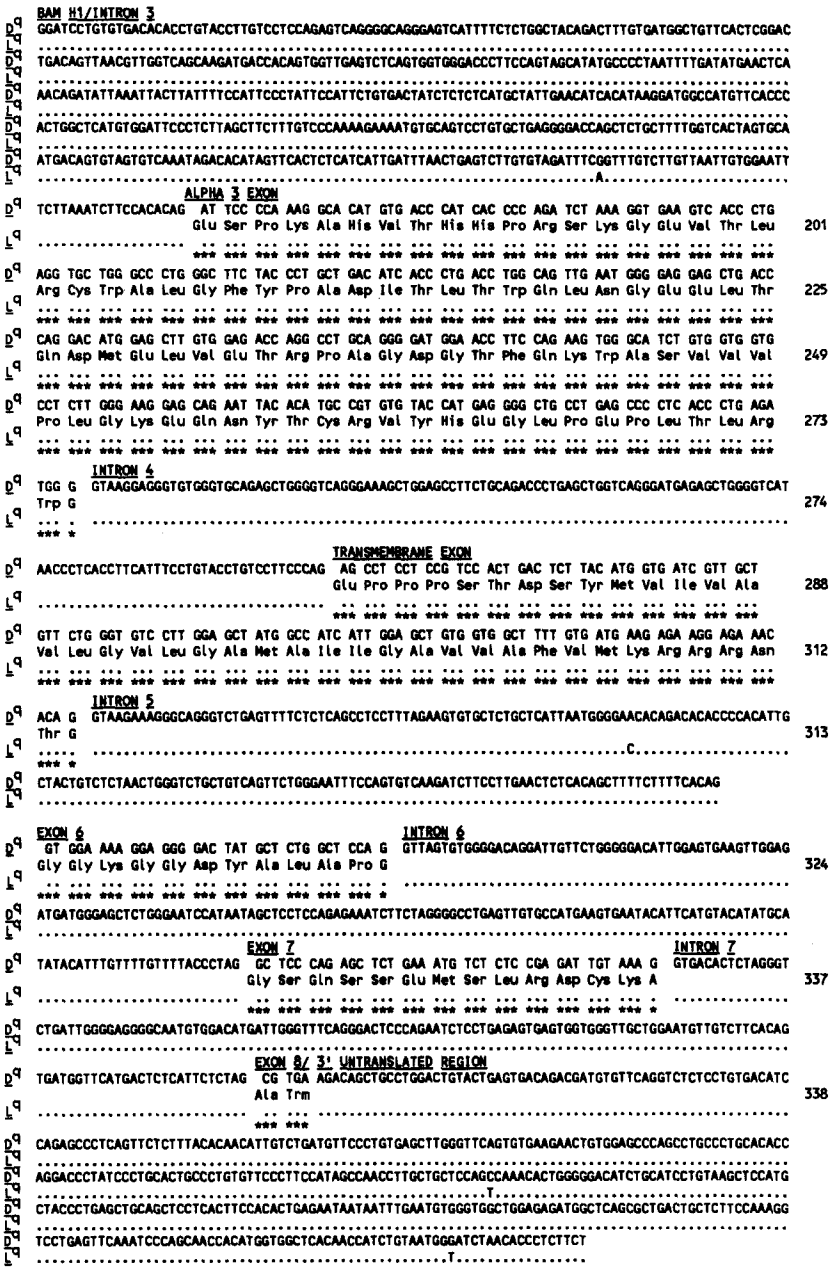

FIGURE 3. DNA sequence and implied amino acid sequence of the 3′ portion of the D^q and L^q genes. Identical sequence is represented by dashes with only the differences between D^q and L^q being noted. Numbering of amino acids is shown. These sequence data have been submitted to the EMBL/GenBank Data Libraries.

```
PROMOTER & 5' UNTRANSLATED REGION
GGATCCTGGGGAACCAAGGAGCTACTCCGCAGGCGCTGGTTATAAAGTCCATGCAGGCCAAGGGTCTCAGATGTCCTGTATTCCAGATGGGGCA
LEADER EXON
ATG GCG CTG CAA AGG CTG CTC CTG CTG CTG GCA GCC GCC CTG ACC CTG ACC AAG ACC CGA GCA G                  64
Met Ala Leu Gln Arg Leu Leu Leu Leu Leu Ala Ala Ala Leu Thr Leu Thr Lys Thr Arg Ala G
INTRON 1
GTGAGTGCGGGGGTCCGCAGGGAAACAGACCTTTCCAAACAGTCTGCGGGGAGGGGCGGGCACAGCACCGGGGAAGCTGCGTGCCCGCGTCGCCCACCGT  164
ACTCTCAGTCTCTTCTCCACCTGCGTCCCGAGCCCCACGCCCTGTTCCCTCCCCGTCCGCGCACCCGCCCAGGGTCCTGGGAGGAGGTCAGGTTCTCACA  264
GCGCGCCGCCCCCAG                                                                                     279
ALPHA 1 EXON
 GC TCA CAC TCG CTG CGG TAT TTC CAC ACC GCC ATG TCC CGG CCG GGC CTT CAG GAG CCC TGG TTC ATC TCT GTC    353
 ly Ser His Ser Leu Arg Tyr Phe His Thr Ala Met Ser Arg Pro Gly Leu Gln Glu Pro Trp Phe Ile Ser Val     25

GGC TAC GTG GAC GAC ACG CAG TTC GTG CGC TTC GAC AGC GAC GCG GAG AAT CCG AGA TAC GAG CCT CGT GTG TCG    428
Gly Tyr Val Asp Asp Thr Gln Phe Val Arg Phe Asp Ser Asp Ala Glu Asn Pro Arg Tyr Glu Pro Arg Val Ser     50

TGG ATG GAG CAA GAG GGG CCG GAG TAT TGG GAG GAG CAG ATA CAG ATC GCC AAG AGC AAT GAG CAG AGA TTC CGA    503
Trp Met Glu Gln Glu Gly Pro Glu Tyr Trp Glu Glu Gln Ile Gln Ile Ala Lys Ser Asn Glu Gln Arg Phe Arg     75

GGG AGC CTG AGG AAC GCG CTG CGC TAC TAC AAC CAG AGC GAG GGC G                                        549
Gly Ser Leu Arg Asn Ala Leu Arg Tyr Tyr Asn Gln Ser Glu Gly G                                         90
INTRON 2
GTGAGTCATGACTCCTCATCTTCCCGGCACCGGGTTGGAGACGTCCCGGATCCGAAGTTTGGGAGCAGAACCGCCCAGAGACTTGTTGCCTTTTCAGTTT  649
GGAGGAGTCCGCGGGAGGGCGGGGCTGACCGCGGGGTCCCGCAG                                                        693
ALPHA 2 EXON
 GC TCT CAC ACG TTC CAG TGG ATG TCT GGC TGT GAC CTC GG- --- --- --- --- C TCC TCC GCG GGT ACA TGC AGT  755
 ly Ser His Thr Phe Gln Trp Met Ser Gly Cys Asp Leu Gl- --- --- --- --- y Ser Ser Ala Gly Thr Cys Ser  111

TCG CTT ATG AAG GCT GCG ATT ACA TCG CCC TGA ACG AAG ACC TGA ACA CGT GAA CTG TGG CGG ACA TGG CGG CGC  830
Ser Leu Met Lys Ala Ala Ile Thr Ser Pro trm                                                          121

AGA TCA CCC GAC ACA ACT GCC AGG AGG CTG GTG CTG CAG AGA AAT AAA TAC AGG GCC TAC CTG GAG GGC ACC TGC  905

TTG GAG AGG CTC CTC AGA TAC CTG CAG CTC CGG AAG GAG ATC GTG CTG CGC ACA G                             960
INTRON 3
GTGCAGGGGCCGCGGGCAGCTCCTCCCTCTGCCCTCGGGCTGGGGCTCAGTCTG                                              1014
```

FIGURE 4. DNA sequence of the 5′ portion of $D4^q$ and its implied amino acid sequence. The CCAAT and TATAA sequences are underlined. The dashes in the α2 exon indicate the locations of the 13-bp deletion that results in a premature termination of the D4 gene product. The location of a 4-bp insertion that creates a potential polyadenylation site is also underlined in the α2 exon. Nucleotide numbering begins at the leader exon.

ences when compared with L^d, whereas the L^q gene has eight nucleotide substitutions resulting in six amino acid differences when compared with L^d (Fig. 5).

Interestingly, the D^q and L^q genes share two short stretches of unique sequences that distinguish them from L^d. These uniquely shared sequences are found at nucleotide positions encoding amino acids 93–96 and 155–157 (Fig. 5). Although these shared sequences could have arisen from gene conversion events between the L^q and D^q genes or from a common donor, this would appear unlikely since it would require multiple independent events. Thus, the unique sequences shared by D^q and L^q genes strongly suggest these genes arose by duplication. In support of this conclusion, the remaining exons (4–8) are identical among D^q, L^q, and L^d. Also, there are common nucleotide differences that both D^q and L^q genes have in introns 1, 3, and 6, in comparison with L^d (reference 17 and Fig. 3). This is in contrast to the BALB/c genes, D^d and L^d, which do not show this high level of homology with each other.

Sequence comparisons of $D4^q$ with D^q and L^q demonstrate that $D4^q$ is not closely related to either, indicating that it is not an L^d-like gene (Fig. 5). While the leader and α1 exons of $D4^q$ appear to contain typical class I sequences, the promoter region and the α2 domain have aberrancies that suggest that the $D4^q$ gene is not functional (Fig. 4). First, the CAAT box sequence (underlined in Fig. 4) is not the typical sequence found among most class I genes. Second, the α2 domain has both a 13-nucleotide deletion and a four-nucleotide insertion. The deletion would cause a frame shift, resulting in a pair of termination codons located 52 and 64 nucleotides 3′ of

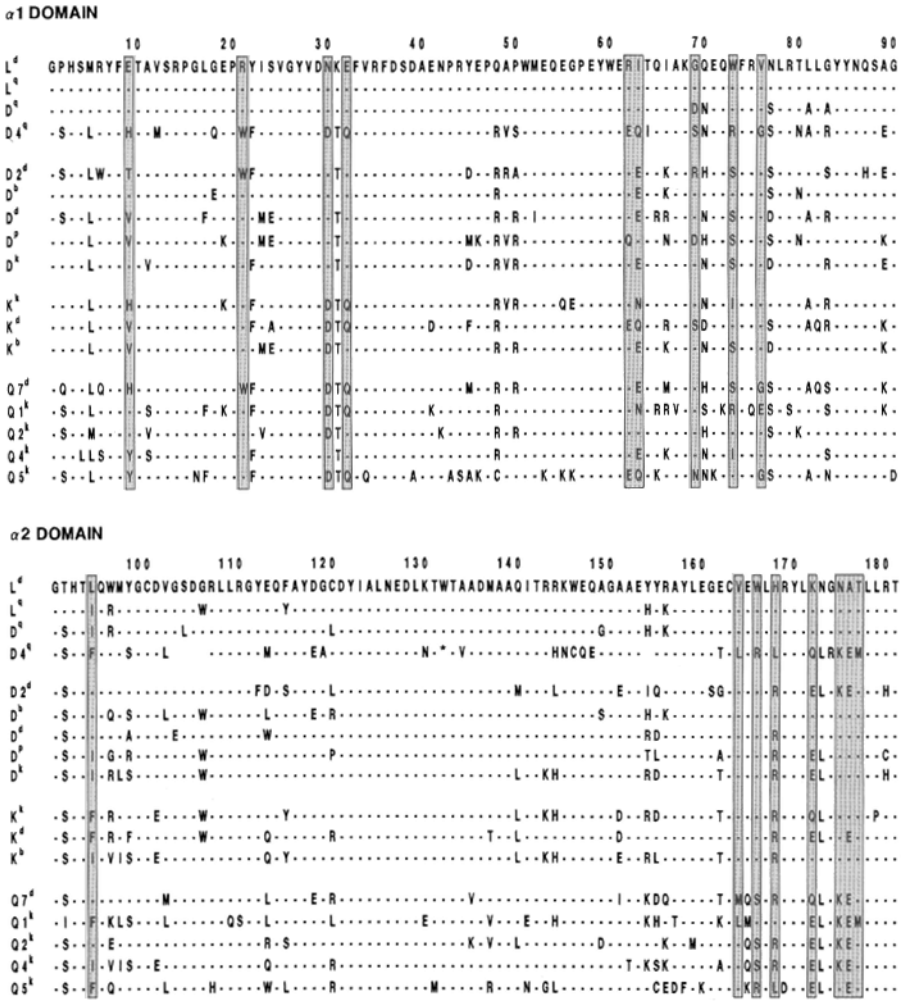

FIGURE 5. Amino acid sequence comparisons of the α1 and α2 domains in class I molecules. Dashes represent identical sequence with L^d. Sequences that are boxed represent some level of common substitutions of $D4^q$ with K/Q region gene products (see references 27, 28, 38–44).

the deletion. In addition, the insertion of four nucleotides creates a frame shift and a potential polyadenylation signal.

Supporting the theory that *D4*^{q} arose by a recombination event between the D and Q region, it can be seen in Fig. 5 that, in an amino acid comparison with other class I molecules, there are individual residues that appear common to some Q and/or K region molecules. The residues are indicated in Fig. 5. Noteworthy are the residues at positions 30, 32, 95, 175, and 176. The amino acids in *D4*^{q} at these positions are found in multiple K/Q region molecules. It was previously shown that position 30 could differentiate D region molecules from K/Q region molecules (25). In addi-

tion, some of these highlighted residues have common substitutions in $D2^d$, which would indicate some relationship between it and K/Q region genes. Thus, the $D4^q$ gene does not appear as closely related to *D* and *L* genes as it does to Q/K region genes. More specifically, the low levels of homology of $D4^q$ with D^q and L^q suggest that it does not represent a recent duplication of a primordial L^d-like gene. The Q region origin of the *D2*, *D3*, and *D4* genes is further strengthened by the results mentioned above, in that both the $D2^q$ and $Q1^q$ genes hybridized strongly to a $D2^d$-derived oligonucleotide probe.

Peptide Binding and TCR Recognition Sites in Class I Molecules. We have analyzed the L^d, L^q, and D^q molecules with regard to the known HLA-A2 crystal structure (3). A number of amino acid residues that are predicted to contact the antigenic peptide are conserved among the L^q, D^q, and L^d molecules. In Table II, the amino acid substitutions among class I molecules compared with L^d are distinguished into those influencing antigenic peptide binding and those influencing the interaction with the TCR. In comparison with L^d, the L^q molecule has two and the D^q molecule has four nonconservative substitutions that influence peptide binding. In fact, if L^q is used as the standard for these comparisons, both D^q and L^d differ from L^q by only two nonconservative substitutions in the peptide-binding site. In contrast, other D region and K region class I molecules range from 7 to 12 nonconservative substitutions when compared with L^d. In addition, fewer substitutions in the sites predicted to interact with the TCR were observed among the L^d-like molecules as compared with the other D region and K region class I molecules. If the results concerning antigen presentation by class II mutant molecules (26) are applicable to class I molecules, then one might predict that the L^d, L^q, and D^q molecules should be able to present a common set of peptides that differ from those presented by other D and K region class I molecules. The ability to present these antigens may be a selective force in maintaining an L^d-like (or L^q-like) structure for these molecules. In strong support of this hypothesis, recent studies of ours have identified peptide ligands shared by various combinations of L^d-like molecules (26a). Importantly, however, the D^q, L^q, and L^d molecules can also be discriminated by their interaction with certain peptide ligands. Thus, this family of L^d-like molecules would appear to be functionally quite similar, but yet, each member also displays subtle distinguishing features.

Analysis of the D Region Organization of Other Haplotypes. Because the D^d and D^q regions were found to have a strikingly similar genetic organization, it was of interest to determine whether other haplotypes shared this same five-gene D region organization. Extensive genomic DNA blot comparisons were made using the aforementioned D region-specific low-copy probes. DNA from various mouse strains representing standard or wild-derived haplotypes were analyzed. Under stringent conditions, probe 50.2A (Fig. 6), isolated from a $D3^d$-containing cosmid clone, 50.2 (12), hybridizes to a single restriction endonuclease fragment from mice containing the D^d, D^q, D^v, and D^r regions, whereas no band was detected from mice containing D^b, D^{dm1}, D^{dm2}, D^k, D^s, D^p, D^{w16}, D^f, and D^j regions. The absence of hybridization of this probe was expected from the D^{dm1}, D^{dm2}, D^b, and D^k regions that are known to contain a single D region gene. A similar genomic analysis was performed using a *D4*-derived probe, 33.3A (see Fig. 1). Again, two fragments were detected from the D^d, D^q, D^r, and D^v regions (data not shown). Thus, these findings would

TABLE II

Amino Acid Substitutions in the α1 and α2 Domains that Influence Antigen Presentation and TCR Recognition

L^d vs.	Peptide*	TCR‡	Other	Total
L^q	2/4§	1/1	1/2	3/6‖
D^q	4/6	2/3	2/5	7/13‖
D^b	8/8	1/2	4/8	12/17‖
D^d	10/13	2/2	6/11	17/25‖
D^k	7/12	3/4	11/16	19/30¶
D^p	12/14	5/5	12/15	27/32¶
K^k	10/14	3/4	12/18	23/34¶
K^b	11/17	2/3	5/12	16/30¶
K^d	8/14	3/4	10/15	21/33

* Amino acid residues proposed to affect the interaction of class I with the peptides found in the binding pocket (3).

‡ Amino acid residues proposed to affect the interaction of class I with the TCR (3).

§ Nonconservative substitutions/total number of substitutions. Conservative amino acid groups are A,G,P,S,T; L,I,V,M; D,E,N,Q; K,R,H; F,Y,W; and C.

‖ Substitutions in one amino acid site that may interact with either the peptide or TCR.

¶ Substitutions in two amino acid sites that may interact with either the peptide or TCR. Amino acid sequences for the comparisons in this table were obtained from references 14, 17, 34–38, 40.

predict that, in addition to the D^q and D^d regions, the D^v and D^r regions share a five-gene organization. This prediction regarding the similarity of d, q, and v haplotypes is consistent with published data by Duran et al. (27), in which a *D3*-associated hybridization probe was used. In contrast to the above results that appear to unambiguously distinguish two *D* region organizations (i.e., a five-gene and a one-gene organization), our results using D region–specific 3′ flanking probes indicated greater complexity. For example, the L^d-derived 3′ flanking probe, L^d 3′Flc, hybridized to two fragments from D^d region DNA, digested with Bam HI and Eco RI (Fig. 6).

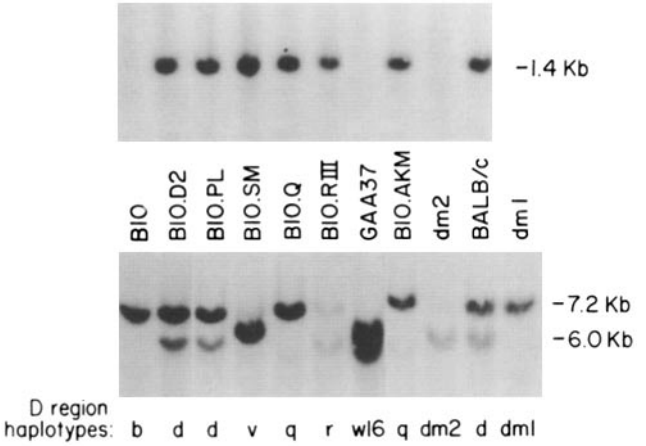

FIGURE 6. DNA blot analyses on various haplotypes using a probe, 50.2A, derived from the $D3^d$ region (*top*) and probe L^d 3′Flc, derived from the 3′ flanking region of L^d (*bottom*). Control lanes dm2 and dm1 identify the D^d and L^d fragments, respectively (*bottom*). The blot shown on top was genomic DNA digested with Bam HI, and the blot on bottom was DNA digested with Bam HI and Eco RI. The D region haplotypes are shown along the bottom. See Fig. 1 for probe location.

Using DNA from the dm1 and dm2 mutants as controls, the L^d- and D^d-associated fragments were identified. Furthermore, only a single band that comigrated with the L^d-specific BALB/c band was detected in DNA from mice containing the D^b region. Thus, this probe appears to be D region specific and capable of determining which other D regions contain *D* and/or *L* genes. However, the hybridization patterns seen with the L^d 3'Flc probe yielded three unexpected findings. First, *D*- and *L*-specific bands were not detected in all haplotypes that tested positive with the D3/D4 probes. For example, only a single band that comigrates with the L^d-specific fragment is detected from the D^q region with this probe; it most likely contains both D^q- and L^q-specific fragments. Second, multiple bands were detected with the L^d 3' Flc probe in some strains that tested negative for the *D3/D4* genes. For example, although D^{w16} DNA was negative using probe 50.2A, two bands hybridized to the L^d 3' Flc probe. Interestingly, these two D^{w16}-associated RFLPs appeared to migrate as a doublet aligning more with the D^d than the L^d gene. The third and related observation was that, using the L^d 3' Flc probe, it is impossible to distinguish between D vs. L alleles. For example, many of the D regions appear to possess unique RFLP patterns that comigrate with neither the L^d nor D^d fragments. Using another flanking probe, II2.20, we also generated data supporting each of the three above-stated observations (data not shown). Collectively, these results strongly suggest that there are several D region organizations, in addition to those containing five genes (d, q) or one gene (b, k). This prediction can only be substantiated by the rigorous molecular characterization of additional independent D regions.

Discussion

The findings reported here demonstrated that the D^q region, like the D^d region, contains five class I genes, and that both regions have very similar restriction maps. We have thus used the same nomenclature for the D^q region-encoded genes as previously used for the D^d region. In a telomeric direction, these genes are D^q, $D2^q$, $D3^q$, $D4^q$, and L^q. Of these genes, we have sequenced the entire D^q and L^q genes, as well as the 5' portion of the $D4^q$ gene. Comparisons of these sequences with pre-

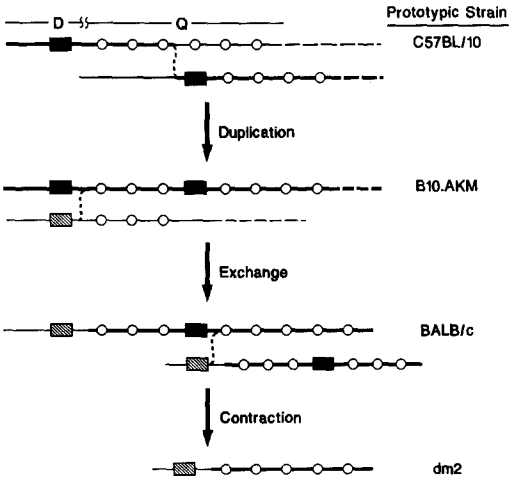

FIGURE 7. Proposed model for the evolution of the D region depicting the continuous expansion and contraction in gene number. The prototypic strains shown were chosen because they contain L^d family genes in their D region, and the dm2 strain was derived directly from the BALB/c strain.

viously reported class I sequences revealed several interesting insights into the genetic evolution of the D^q region. The D^q and L^q genes were found to be strikingly similar, including an identical 3′ half. Furthermore, both the L^q and D^q genes are structurally very similar to the L^d and D^b genes, yet all of the aforementioned genes are quite distinct from the D^d gene. These findings are consistent with the evolutionary model shown in Fig. 7. According to this model, a primordial D region contained a single D region gene. For example, the D^b region of strains such as C57BL/10 (B10) would represent such a prototype containing a single L^d-like class I gene. This ancestral D^b-like gene could have duplicated to evolve into a D region containing two D^b-like genes, namely *D* and *L*. The D^q region provides evidence for such a duplication event. Not only are the D^q and L^q genes very similar to each other, but they share unique sequences that distinguish them from other class I genes. Furthermore, our data suggest that this putative duplication probably resulted from the misalignment of the ancestral *D* gene with a Q region gene. Such an unequal crossover event, as shown in Fig. 7, would result in the introduction of three *Q*-like genes intercalated between two *D*-like genes. This model was initially proposed on the basis of restriction maps, and the crosshybridization of low-copy probes to the D and Q regions (12). A subsequent finding consistent with this model is that the *D2*, *D3*, and *D4* genes lack a *B2* sine Alu repeat found in the 3′ UT region of all *D* and *L* genes investigated (27). The derivation of the *D2-D4* genes from *Q* genes is strongly supported by two findings reported here. First, the genomic sequence of the $D4^q$ gene shows both specific and overall similarities to published Q and K region class I genes. The similarities of the latter genes with each other have been interpreted as evidence that the K region evolved from a translocation of Q region genes (6). In any case, the $D4^q$ gene is not particularly similar to other D region genes, including its neighbors D^q and L^q. Thus, the $D4^q$ gene most probably did not result from the duplication of an ancestral D region gene. The second observation that suggests the *D2*, *D3*, and *D4* genes originated from the Q region is that the oligonucleotide probe specific for the transmembrane segment of the $D2^d$ gene was found to hybridize to both the $D2^q$ and the $Q1^q$ genes. This latter gene was also isolated from the B10.AKM cosmid library and was provisionally mapped immediately telomeric of the L^q gene. Thus, these findings support the proposed model that the *D2*, *D3*, and *D4* genes evolved from ancestral *Q* genes introduced into the D region by an unequal crossing over resulting from the misalignment of a primordial *D* gene with a *Q* gene. Given the validity of this model, it could further be proposed that the D^d region evolved from the D^q region by an homologous recombination event substituting a D^d-like gene with the D^q gene. Such an evolutionary scheme is supported by the striking similarity of the D^q, L^q, L^d, and D^b genes, in contrast to the dissimilarity of the D^d gene. As will be discussed below, there are now three documented cases of intra-D region recombination (one homologous, two nonhomologous) thus validating the feasibility of the proposed model.

Unfortunately, characterization of the *D2*, *D3*, and *D4* genes and their potential products is incomplete, and what relevant information is available is from different haplotypes. The $D2^d$ gene has recently been sequenced and appears to be a functional gene; however, a $D2^d$ product has not yet been detected (28). In contrast to these findings, we present evidence here that the $D4^q$ gene could not encode a typical class I molecule. Even if there is a $D4^q$ product, it would represent only the

5′ half of the gene. Further, the $D2^d$ and $D4^q$ genes are not particularly similar to each other or to other D region genes. As discussed above, there is considerable circumstantial evidence that the *D2*, *D3*, and *D4* genes originated from Q region genes, perhaps *Q1-Q3*. However, Goodenow and associates (29) recently published the sequence of the *Q* genes of the k haplotype, and neither $D2^d$ nor $D4^q$ bear striking similarity to any of the Q^k genes. This lack of similarity could reflect either: (*a*) that the unequal crossing over that introduced *D2*, *D3*, and *D4* genes into the D region occurred sufficiently long ago to permit considerable divergence; or (*b*) there is considerable haplotype-specific polymorphism among the *Q* genes, and the *Q* genes that duplicated to generate *D2*, *D3*, and *D4* genes have yet to be cloned and sequenced. In support of this second alternative, the restriction maps of the Q regions in the d, b, and k haplotype differ considerably (6, 12). To determine the polymorphic nature of *Q* genes and potentially identify recent relatives of the *D2*, *D3*, and *D4* genes, we are cloning *Q* genes from our B10.AKM library.

In addition to the studies reported here of the D^q region, there are three other examples of intra-D region recombination. Both the dm1 and dm2 mutations involved the deletion of multiple genes within the D region of the H-2^d haplotype. The dm1 mutant contains a single chimeric gene, D^{dm1}, that is identical to D^d in its 5′ half and identical to L^d in its 3′ half (15). The dm2 mutant contains only the D^d gene (16). Thus, both the dm1 and dm2 mutations reduced the D region gene number from five to one, and both probably resulted from misalignments of the D^d and L^d genes leading to an unequal crossover event (15, 16). A third example of an intra-D region recombination was described in the B10.RQDB strain (30). This strain resulted from a recombination event between the D^d and D^b regions in a heterozygous mouse, such that the L^d gene was replaced by the highly homologous D^b gene. The resulting D region of B10.RQDB contains the D^d, $D2^d$, $D3^d$, $D4^d$, and D^b genes. Therefore, this recombination event introduced the D^b gene, which is normally a single D region gene, into a five-gene D region organization. The above studies of B10.AKM, dm1, dm2, and B10.RQDB all provide evidence that the D region is in a dynamic state of genetic expansion and contraction. Given these results, it will be very interesting to determine whether there are additional D region gene organizations besides the one-gene (d, k) and the five-gene (d, q) organizations. Although previous genomic DNA blot analyses indicated there may be only two D region gene organizations (12), our data suggest there could be several different organizations. For example, DNA from strains such as B10.GAA37 (D^{w16}) did not hybridize with low-copy probes to *D2*, *D3*, or *D4* genes, yet appeared to have multiple genes using D region-specific 3′ flanking probes. Preliminary molecular studies of B10.GAA37 (D. Lee, manuscript in preparation) support the results of genomic DNA analyses presented here, and those previously described by serological and peptide mapping studies (9). Therefore, B10.GAA37 appears to possess a D region organization distinct from the five-gene and one-gene organization described here.

This haplotype disparity in gene number appears to be a unique property of the D region and does not apply to the K region. In haplotypes studied thus far, the K region contains two genes, only one of which is functional. Interestingly, the Q region has been found to display extensive haplotype diversity with gene number varying from 1 to 10 (5, 6, 29). Similarly, haplotype diversity in gene number appears to be a trait of the Tla region (2, 5, 6). It is very intriguing that the centromeric

K region should remain static, whereas the more telomeric D, Q, and Tla regions appear to have undergone genetic expansion and contraction.

The results reported here are also quite significant when compared with recent studies of the human HLA A, B, and C loci. It has widely been speculated that the *L* gene of the mouse is the homologue of the *C* gene of humans. This conclusion is supported by much correlative data, including the fact that both *L* and *C* genes appear less polymorphic and include null phenotypes (31). Furthermore, the respective products of C and L loci appear to have a weaker affinity for β_2m and a lower surface expression than other class I molecules (32–34). However, the data presented here suggest that these above correlations are merely coincidental and thus refute the homologous nature of the *L* and *C* genes. Whereas the D/L regions of the mouse appear to be in a dynamic state of gene expansion and contraction, the HLA-B/C regions appear fixed in terms of gene number. As mentioned above, gene deletions (dm1, dm2) or gene duplications (D^q L^q) cause the disparity in *D/L* gene number in the mouse. By contrast, the C blanks in humans have recently been shown to result from less antigenic products at this locus and not HLA-C deletions (35, 36). The consequence of this is that structural comparisons of HLA class I molecules clearly indicate a "C-ness," as well as a "B-ness" and "A-ness" when comparing allelic products (37). By contrast, in the mouse, it is not even clear which *D/L* genes are alleles. For example, in haplotypes with a single D region gene, such as D^k or D^b, it is unclear from their restriction maps or their sequence whether to align them with *D* or *L* genes. Indeed, the 5′ flanking restriction maps D^k and D^b look more like D^d, whereas the 3′ flanking restriction maps of D^k and D^b look more like that of the L^d gene. In another example, the D^q gene is strikingly similar to the L^d gene, yet clearly aligns with the D^d locus. Thus, the term allele is not applicable to *D* or *L* genes in mice, but is to the *A*, *B*, and *C* genes in humans. Furthermore, the inability to define D/L alleles in the mouse precludes relevant comparisons, such as levels of polymorphism, β_2m associations, or cell surface expression. For example, the D^k molecule, like L^d, has a weak β_2m association and lower cell surface expression (38), yet there is no criteria by which D^k and L^d can be considered alleles. Parham et al. (37) suggested that the lack of locus-specific sequences in mouse reflects that their polymorphism resulted from inter-loci gene conversion events, whereas in human intra-loci conversion events appear to have generated the polymorphism. However, our data would suggest that the genetic expansion/contraction of the class I genes at least in the D region is the cause of the lack of "D-ness" and "L-ness." In any case, by several criteria, the genetic mechanisms generating polymorphism in humans and mice appear to be significantly different.

Summary

To resolve issues regarding the evolution of D region class I MHC genes and their relationship to other class I–encoding regions of the mouse, as well as man, we characterized the class I genes from the D^q region of the B10.AKM mouse strain. The D^q region was selected because it was known to express multiple gene products, yet two of the products previously characterized have structural features in common with the L^d molecule. Since DNA hybridization data defined similarities between the D^d and D^q regions, we used low-copy genomic or oligonucleotide probes de-

rived from the D^d region of BALB/c (H-2^d) to screen a B10.AKM cosmid library. Cosmid clones containing D^q, $D2^q$, $D3^q$, $D4^q$, L^q, and $Q1^q$ genes have been isolated and aligned with the corresponding genes of the BALB/c MHC, thus demonstrating a similar gene organization. The two classical transplantation genes, D^q and L^q, were found to be strikingly similar to each other such that exons 1–3 of D^q and L^q are ~97% homologous, and exons 4–8 are identical. Furthermore, the implied amino acid sequences of both L^q and D^q molecules show considerable homology to L^d, particularly in regions presumed to be involved in ligand binding. These comparisons suggest not only that the D^q and L^q genes arose from the duplication of an L^d-like progenitor, but also that there is a selective advantage for the maintenance of an L^d-like structure. In addition, the 5′ portion of the $D4^q$ gene was sequenced and found to have a 13-bp deletion and a 4-bp insertion within the $\alpha 2$ exon. These result in a frame shift that creates a premature termination codon and potential polyadenylation site, respectively. Thus, $D4^q$ does not encode a typical class I molecule. Sequence comparisons suggest that the $D4^q$ gene did not arise from a duplication event involving an L^d-like gene such as D^q and L^q. Interestingly, the $D4^q$ molecule, if produced, would have amino acid residues in common with K and/or Q molecules that differ from those observed in D/L molecules. These findings, in conjunction with hybridization data, provide evidence that the *D2*, *D3*, and *D4* genes were derived from *Q* genes by an unequal crossover event. Additional hybridization data using low-copy D region probes suggest that several different D region gene organizations exist among mice of different haplotypes. These and other recent molecular studies provide multiple examples of expansion and contraction of the class I genes in the D region. This feature of the H-2 *D/L* genes distinguishes them from the HLA *B/C* genes, suggesting significant differences between the evolution of the MHC in man and mouse.

We thank Dr. David Chaplin for scientific advice and cosmid sequencing primers; Bill Walsh and Shawn McVey for technical help; Dr. Steven Hunt for oligonucleotide probes; Drs. Janet Connolly and Nancy Johnson for critical review of manuscript; and Karen Perks for her excellent secretarial work.

Received for publication 7 December 1989 and in revised form 9 February 1990.

References

1. Townsend, A., J. Rothbard, F. Gotch, G. Bahadur, D. Wraith, and A. McMichael. 1986. The epitopes of influenza nucleoprotein recognition by cytotoxic T lymphocytes can be defined with short synthetic peptides. *Cell.* 44:959.
2. Flaherty, L. 1981. The Tla region antigens. *In* The Role of the Major Histocompatibility Complex in Immunobiology. M. E. Dorf, editor. Garland, New York. 33–58.
3. Bjorkman, P. J., M. A. Saper, B. Samraoui, W. S. Bennett, J. L. Strominger, and D. C. Wiley. 1987. The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature (Lond.).* 329:512.
4. Nathenson, S. G., J. Geliebter, G. M. Pfaffenbach, and R. A. Zeff. 1986. Murine major histocompatibility complex class I mutants: molecular analysis and structure-function implications. *Annu. Rev. Immunol.* 4:471.
5. Steinmetz, M., D. Stephan, and K. F. Lindahl. 1986. Genetic organization and recombinational hotspots in the murine major histocompatibility complex. *Cell.* 44:895.

6. 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.
7. Hansen, T. H., K. Ozato, M. R. Melino, J. E. Coligan, T. J. Kindt, T. J. Jandinski, and D. H. Sachs. 1981. Immunochemical evidence in two haplotypes for at least three D region-encoded molecules. *J. Immunol.* 126:1713.
8. Lillehoj, E. P., T. H. Hansen, D. H. Sachs, and J. E. Coligan. 1984. Primary structural evidence that the H-2D^q region encodes at least three distinct gene products: D^q, L^q, and R^q. *Proc. Natl. Acad. Sci. USA.* 81:2499.
9. Lillehoj, E. P., W. D. Walsh, T. Potter, D. R. Lee, J. E. Coligan, and T. H. Hansen. 1984. Chemical and serological definition of two unique D region-encoded molecules in the wild-derived mouse strain B10.GAA37. *J. Immunol.* 133:3138.
10. Lillehoj, E. P., N. B. Myers, D. R. Lee, T. H. Hansen, and J. E. Coligan. 1985. Structural definition of a family of L^d-like molecules distributed among four of seven haplotypes compared. *J. Immunol.* 135:1271.
11. Lee, D. R., E. P. Lillehoj, N. B. Myers, J. E. Coligan, and T. H. Hansen. 1986. Structural analyses define an additional H-2D region class I antigen that is expressed by a variant mouse strain. *Mol. Immunol.* 23:1093.
12. Stephan, D., H. Sun, K. F. Lindahl, E. Meyer, G. Hammerling, L. Hood, and M. Steinmetz. 1986. Organization and evolution of D region class I genes in the mouse major histocompatibility complex. *J. Exp. Med.* 163:1227.
13. Watts, S., J. M. Vogel, W. D. Harriman, T. Itoh, H. J. Stauss, and R. S. Goodenow. 1987. DNA sequence analysis of the C3H H-2K^k and H-2D^k loci: evolutionary relationships to H-2 genes from four other mouse strains. *J. Immunol.* 139:3878.
14. Schepart, B. S., H. Takahashi, K. M. Conrad, P. Murray, K. Ozato, E. Appella, and J. A. Frelinger. 1986. The nucleotide sequence and comparative analysis of the H-2D^p class I gene. *J. Immunol.* 136:3489.
15. Sun, Y. H., R. S. Goodenow, and L. Hood. 1985. Molecular basis of the dm1 mutation in the major histocompatibility complex of the mouse: a D/L hybrid gene. *J. Exp. Med.* 162:1588.
16. Rubocki, R. J., T. H. Hansen, and D. R. Lee. 1986. Molecular studies of murine mutant BALB/c-H-2^{dm2} define a deletion of several class I genes including the entire H-2L^d gene. *Proc. Natl. Acad. Sci. USA.* 83:9606.
17. Lee, D. R., R. J. Rubocki, W. R. Lie, and T. H. Hansen. 1988. The murine MHC class I genes, *H-2D^q* and *H-2L^q*, are strikingly homologous to each other, *H-2L^d*, and two genes reported to encode tumor-specific antigens. *J. Exp. Med.* 168:1719.
18. Chang, H., E. Omitrovsky, P. A. Hieter, K. Mitchell, P. Leder, L. Turoczi, I. R. Kirsch, and G. F. Hollis. 1986. Identification of three new Ig λ-like genes in man. *J. Exp. Med.* 163:425.
19. Hanahan, D., and M. Meselson. 1983. Plasmid screening at high colony density. *Methods Enzymol.* 100:333.
20. Kress, M., Y. Barra, J. G. Seidman, G. Khoury, and G. Jay. 1984. Functional insertion of an Alu type 2 (B2 SINE) repetitive sequence in murine class I genes. *Science (Wash. DC).* 226:974.
21. Feinburg, A. P., and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132:6.
22. Brorson, K. A., S. W. Hunt III, T. Hunkapiller, Y. H. Sun, H. Cheroutre, D. A. Nickerson, and L. Hood. 1989. Comparison of exon 5 sequences from 35 class I genes of the BALB/c mouse. *J. Exp. Med.* 170:1837.
23. Graham, M. Y., T. Otani, I. Boime, M. V. Olson, G. S. Carle, and D. D. Chaplin. 1987. Cosmid mapping of the human chorionic gonadotropin β subunit genes by field inversion gel electrophoresis. *Nucleic Acids Res.* 15:4437.

24. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463.
25. Klein, J., and F. Figueroa. 1986. The evolution of class I MHC genes. *Immunol. Today.* 7:41.
26. Reske-Kunz, A. B., D. Landais, J. Peccoud, C. Beneist, and D. Mathis. 1989. Functional sites on the Aα-chain. Polymorphic residues involved in antigen presentation to insulin-specific, A^b_α:A^k_β-restricted T cells. *J. Immunol.* 143:1472.
26a. Lie, W.-R., N. B. Myers, J. Gorka, R. J. Rubocki, J. M. Connolly, and T. H. Hansen. 1990. Peptide ligand-induced conformation and surface expression of the L^d class I MHC molecule. *Nature (Lond.).* 344:439.
27. Duran, L. W., R. M. Horton, C. W. Birschbach, A. Chang-Miller, and L. R. Pease. 1989. Structural relationships among the H-2D regions of murine MHC haplotypes. *J. Immunol.* 142:288.
28. Hedley, M. L., S. W. Hunt III, K. A. Brorson, J. S. Andris, L. Hood, J. Forman, and P. W. Tucker. 1989. Analysis of $D2^d$ a D region class I gene. *Immunogenetics.* 29:359.
29. Watts, S., A. C. Davis, B. Gaut, C. Wheeler, L. Hill, and R. S. Goodenow. 1989. Organization and structure of the Qa genes of the major histocompatibility complex of the C3H mouse: implications for Qa function and class I evolution. *EMBO (Eur. Mol. Biol. Organ.) J.* 8:1749.
30. Duran, L. W., J. C. Zeller, J. K. Lundy, A. Chang-Miller, C. J. Kroc, C. S. David, and L. R. Pease. 1987. Genetic analysis of the H-2D region using a new intra-D region recombinant mouse strain. *J. Immunol.* 139:2818.
31. Aizawa, M. 1986. Antigens and gene frequencies of ethnic group. *In* HLA In Asia-Oceania. M. Aizawa, T. Natori, A. Wakisaka, and Y. Konoeda, editors. Hokkaido University Press, Sapporo, Japan. 1079–1091.
32. Gusson, D., R. S. Rein, I. Mejer, W. de Hoog, G. H. A. Seemann, F. M. Hochstenbach, and H. L. Ploegh. 1987. Isolation, expression, and the primary structure of HLA-Cw1 and HLA-Cw2 genes: evolutionary aspects. *Immunogenetics.* 25:313.
33. 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.
34. Snary, D., C. J. Barnstable, W. F. Bodmer, and M. J. Crumpton. 1977. Molecular structure of human histocompatibility antigens: the HLA-C series. *Eur. J. Immunol.* 7:580.
35. Takiguchi, M., I. Nishimura, H. Hayashi, S. Karaki, A. Kariyone, and K. Kano. 1989. The structure and expression of genes encoding serologically undetected HLA-C locus antigens. *J. Immunol.* 143:1372.
36. Tibensky, D., F. Decary, and T. L. Delovitch. 1988. HLA-C genes are transcribed in HLA-C blank individuals. *Immunogenetics.* 27:220.
37. Parham, P., C. E. Loman, D. A. Lawlor, J. P. Ways, N. Holmes, H. L. Coppin, R. D. Salter, A. W. Wan, and P. D. Ennis. 1988. Nature of polymorphism in HLA-A, -B, and -C molecules. *Proc. Natl. Acad. Sci. USA.* 85:4005.
38. Williams, D. B., S. J. Swiedler, and G. W. Hart. 1985. Intracellular transport of membrane glycoproteins: two closely related histocompatibility antigens differ in their rates of transit to the cell surface. *J. Cell Biol.* 101:725.
39. 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 antigens. *Science (Wash. DC).* 215:679.
40. Watts, S., J. M. Vogel, W. D. Harriman, T. Itoh, H. J. Strauss, and R. S. Goodenow. 1987. DNA sequence analysis of the C3H H-$2K^k$ and $H2D^k$ loci: evolutionary relationships to H-2 genes from four other mouse strains. *J. Immunol.* 139:3878.
41. Sher, B. T., R. Nairn, J. E. Coligan, and L. Hood. 1985. DNA sequence of the mouse H-$2D^d$ transplantation antigen gene. *Proc. Natl. Acad. Sci. USA.* 82:1175.

42. Arnold, B., H.-G. Burgert, A. L. Archibald, and S. Kvist. 1984. Complete nucleotide sequence of the murine H-2K^k gene: comparison of three H-2K locus alleles. *Nucleic Acids Res.* 24:9473.
43. 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.
44. Steinmetz, M., K. W. Moore, J. G. Frelinger, B. T. Sher, F.-W. Shen, E. A. Boyse, and L. A. Hood. 1981. Pseudogene homologous to mouse transplantation antigens: transplantation antigens are encoded by eight exons that correlate with protein domain. *Cell.* 25:683.
45. 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.