

# A Nonpolymorphic Class I Gene in the Murine Major Histocompatibility Complex

Andrew L. Mellor,\* Elisabeth H. Weiss,\*  
Michel Kress,† Gilbert Jay,† and Richard A. Flavell\*

\* Biogen Research Corporation  
14 Cambridge Center  
Cambridge, Massachusetts 02142

† Laboratory of Molecular Virology  
National Cancer Institute  
National Institutes of Health  
Bethesda, Maryland 20205

## Summary

**DNA sequence analysis of a class I gene (Q10), which maps to the Qa2,3 locus in the C57BL/10 (H-2<sup>b</sup> haplotype) mouse, reveals that it is almost identical to a cDNA clone (pH16) isolated from a SWR/J (H-2<sup>a</sup> haplotype) mouse liver cDNA library. Exon 5, in particular, has an unusual structure such that a polypeptide product is unlikely to be anchored in the cell membrane. Our findings suggest that the two sequences are derived from allelic class I genes, which are nonpolymorphic, in contrast to H-2K allelic sequences from the same mice, and they may encode liver-specific polypeptides of unknown function. Our previous studies indicate that the Q10 gene is a potential donor gene for the generation of mutations at the H-2K locus by inter-gene transfer of genetic information. Thus the lack of polymorphism in class I genes at the Q10 locus implies either that they are not recipients for such exchanges or that selective pressure prevents the accumulation of mutations in genes at this locus.**

## Introduction

The major histocompatibility complex (MHC) of mouse (H-2 complex) is a large multigene complex that encodes several cell-surface antigens involved in the recognition of and response to foreign antigens by the immune system (Klein, 1975). Class I polypeptides (Ploegh et al., 1981; Coligan et al., 1981) are encoded at four loci within the MHC (H-2K, H-2D, Qa2,3, and TL) as shown by serological and genetic analysis of class I antigens that are expressed on all tissue types (H-2K and H-2D,L antigens) and of biochemically similar antigens that are only expressed on certain lymphoid cells (Qa and TL antigens). One striking characteristic of the class I antigens expressed at the H-2K,D and L loci is their high polymorphism (Klein, 1975), although there is some evidence that the class I antigens that map to the Qa2,3 and TL loci exhibit a much lower level of polymorphism among mice (Michaelson et al., 1983). In total, between six and nine class I polypeptides have been distinguished. This estimate of the number of class I polypeptides contrasts with the numbers of different class I-related gene sequences that have been isolated by molecular cloning. For example, we have isolated 26 dif-

ferent class I-related gene sequences from a cosmid library containing C57BL/10 (H-2<sup>b</sup>) spleen DNA (Golden et al., unpublished data), and 35 gene sequences have been isolated from a Balb/c (H-2<sup>d</sup>) sperm library (Steinmetz et al., 1982; Winoto et al., 1983). The large discrepancy between the number of phenotypically characterized class I polypeptides and the number of class I gene sequences may be explained in two ways: the additional gene sequences that hybridize to the nucleic acid probes are pseudogenes—i.e., they cannot encode a class I cell-surface antigen due to a particular molecular defect; and the extra class I genes encode antigens that have not been, or cannot be, detected. This would occur if the antigens they encoded were not polymorphic between different mouse inbred strains such that alloantisera could not be generated. DNA sequence analysis of cloned class I genes that have no known protein coding function will help distinguish between these two alternatives. One indication that some of the additional class I genes in the Balb/c mouse may express class I polypeptides has been reported by Goodenow et al. (1982), who introduced cloned class I genes into L-cells by DNA-mediated gene transfer and observed increased levels of  $\beta_2$  microglobulin expression in transformed cells with several genes mapping to the Qa and TL loci.

Recently, we have reported that a class I gene located in the Qa2,3 region, which has no known protein-coding function, is a potential donor for a gene conversion event that we believe led to the bml mutation in the H-2K<sup>b</sup> gene (Mellor et al., 1983; we will refer to this gene throughout this report as Q10). We have now obtained DNA sequence for all of the potential exons in this gene. Our results demonstrate that the Q10 gene could encode a functional class I polypeptide but, due to a frame shift in the exon that encodes the transmembrane domain (exon 5), this protein is unlikely to be present in the cell membrane. Similar changes in this domain have been reported for a liver-specific cDNA clone obtained from SWR/J (H-2<sup>a</sup>) mice (Cosman et al., 1982a, 1982b; Kress et al., 1983a). Comparison of this cDNA sequence with the Q10 gene (H-2<sup>b</sup>) sequence reveals that the two sequences are remarkably conserved between the two mice (only seven differences were found in 1368 compared bases), suggesting that they are probably alleles and, furthermore, that these alleles are not polymorphic in the two mice. In contrast, a parallel comparison of H-2K sequences from the same mice reveals characteristically high polymorphism in these alleles. We conclude that the lack of polymorphism in the Q10 gene locus is due to fundamental differences in the evolution of class I genes at the H-2 or Q10 loci.

## Results

### The Q10 Gene

Recently (Mellor et al., 1983), we have reported that the Q10 gene, which was isolated from a C57BL/10 (H-2<sup>b</sup> haplotype) cosmid library, is a potential donor gene for the

bml gene conversion event in the C57BL mouse since it has the bml-specific nucleotide sequence in exon 3, that encodes the second polymorphic protein domain ( $\alpha_2$ ). This gene is located in a large region of DNA (~200 kb), defined by many overlapping cosmids, which contains 11 closely linked class I genes arranged in a head-to-tail tandem array. Furthermore, we have shown that this cluster of class I genes is located in the Qa2,3 region, distal to the H-2D end of the H-2 complex on chromosome 17. Further details of this cluster and of the position of the Q10 gene in this cluster will be published shortly (Golden et al., unpublished data). We decided to obtain the complete sequence of the Q10 gene because of its potential function as a donor for gene-conversion events that lead to polymorphism in the H-2K locus and to find out if this gene, which maps outside the classical H-2 complex, has a gene structure similar to those of the H-2K, D, L genes. Furthermore, we expected to learn whether the Q10 gene could encode a class I polypeptide.

A complete copy of the Q10 gene is present in cosmid B1.30. We established, using appropriate 5' and 3' gene probes, that the entire gene sequence is present in three Bam HI restriction enzyme fragments (of sizes 0.7, 1.3, and 3.0 kb) in this cosmid. All three Bam HI fragments were subcloned into the vector pBR327 and restriction maps for each fragment were deduced. DNA sequence analysis was carried out according to the scheme outlined in Figure 1a using the methods described in Experimental Procedures. The sequence of the three Bam HI fragments is shown in Figure 2.

### The Q10 Gene Is Allelic to an Unusual cDNA Sequence

When we compared the sequence shown in Figure 2 with all available class I DNA sequences we found that it is almost identical with the sequence of a cDNA clone (pH16) that was isolated from a SWR/J (H-2<sup>a</sup> haplotype) liver cDNA library (Cosman et al., 1982a, 1982b; Kress et al., 1983a). In fact, we detected only seven base differences out of 1368 (i.e., 99.5% sequence conservation) between the Q10 and the pH16 (and an unpublished sequence obtained from cDNA clone pH105 that overlaps with, and extends, the pH16 sequence) cDNA sequences. This is a surprisingly high homology to find between two class I sequences from mice carrying different H-2 haplotypes, especially since published exon 2 and 3 sequences (which encode the most polymorphic domains in a class I polypeptide) are no more than 90% homologous, even among alleles (see for example Weiss et al., 1983a; Kress et al., 1983a). In addition, the sequence of exon 5, which encodes the transmembrane domain of class I antigens, is unusual in both sequences in that there is: a 13 base deletion that introduces a frame shift and leads to termination of translation downstream in the same exon; and several base changes, that introduce polar amino acid residues into the transmembrane domain when compared to other class I gene sequences (Cosman et al., 1982a). Cosman et al. (1982b) have suggested that these unusual structural characteristics may prevent a polypeptide product, encoded by the mRNA from which clone pH16 is derived, from being inserted into the cell membrane. In

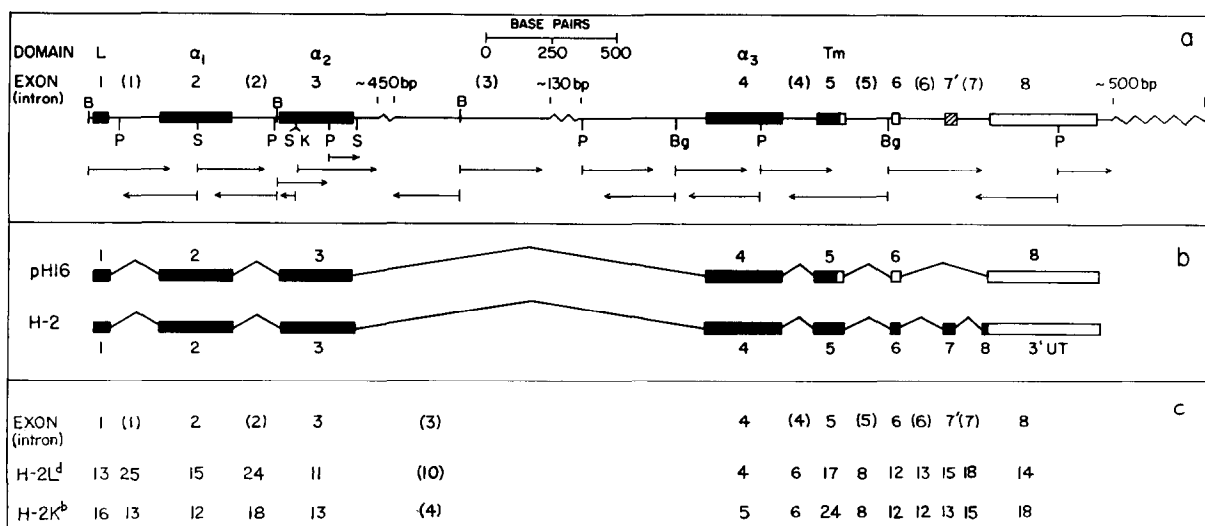


Figure 1. Schematic Diagram of the Q10 Gene, Showing Sequence Strategy and Comparisons with Other Class I Genes

(a) Restriction map of the Q10 gene and sequence strategy. B = Bam HI; Bg = Bgl II; K = Kpn I; S = Sac II; P = Pst I. The seven exon sequences homologous to pH16 cDNA (see text and b) are depicted as thick boxes. The region designated as 7' is not present in mRNA transcripts of the Q10 gene (see below). Closed boxes indicate translated sequences, open boxes indicate untranslated mRNA sequences. Gaps (not to scale) in the sequence are shown as wavy lines. (b) Different splicing patterns in the pH16 cDNA clone (upper line, Kress et al., 1983a) or in a prototype (for example, H-2K<sup>b</sup> and H-2L<sup>d</sup>) H-2 class I gene (see for example Weiss et al., 1983a). In each case, exons are numbered from left to right (5'→3'). The 3' exon of pH16 (and of the Q10 gene) is homologous to prototype exon 8 sequences and, for clarity, we have designated this as exon 8. (c) Percent divergence of Q10 intron (brackets) and exon sequences when compared to either H-2L<sup>d</sup> (Moore et al., 1982) or H-2K<sup>b</sup> (Weiss et al., 1983a) gene sequences. Figures for intron 3 (in brackets) show the diversity in the 5' 90 bases of this region, the maximum available length for comparison.

```

MetAlaProArgThrLeuLeuLeuLeuLeuAlaAlaAlaLeuAlaProThrGlnThrGlnAlaG
GGATCCCAGATGGGGGCGATGGCCCGCGACGCTGCTCCTGCTGCTGGCGGCTGCCCTGGCCCCGACCCAGACCAGGCAGGTGAGTGGCGGGTGGGGAGGGAAACAGCTTCGAGAG- 120
AGGGCGGGGGGGCCAGCAGGAAAGCGGCCCTCCCGGCTCGCCACCGGACCCCTCGCCCTCTTTACCCAGAGCGCCCGGCGCTGCTCCCTCCCGCCCGCCGACCCCGGGGTCC- 240
1ySerHisSerMetArgTyrPheGluThrSerValSerArgProGlyLeuGlyGluProArgPheIleValGlyTyrVa
GGAGAGGAGGTCGCGGCTCACCGTCCAGCAGGCTACACCTCCAGAGGTATTTCGAAACCTCGCTTCCCGCCGGGCCITGGGGAGCCCGGTTCAATATTGTCGGTTACGT- 360
lAspAspThrGlnPheValArgPheAspSerAspAlaGluThrProArgMetGluProArgAlaProTrpMetGluGlnGluGlyProGluTyrTrpGluArgGluThrGlnArgAlaLy
GGAGCAGCAGCAGTTTCGCGCTTCCAGCGCACGGGAGACTCAGAGGATGGAGCGCGCGCCCTTGGATGGACGAGAGGGGGGAGGTATTGGGAGCGGGACACAGAGCCCAA- 480
sGlyAsnGluGlnSerPheHisValSerLeuArgThrLeuLeuGlyHisTyrAsnGlnSerGluSerG
GGGCAATGAGCAGAGTTCCATGTAGCCCTGAGGACCTGCTCGGCEACTACAACCAGAGCGAGGCGGTGAGTGACCCTGGAGGTCAGACCCTCCACGTCCTCCAAACCGAGGCGGTG- 600
ACGTCCTCCGGTCAAAAGTCCAGGTTTCGGGACGAGAATGACCCAGGACCCGATTCCTTTTCAGTTCAGGGTCCCGCAGGGGGGAGGGGGGGGGGGAGGGGGGCTGACTCGAGGATC- 720
1ySerHisThrIleGlnTrpMetTyrGlyCysLysValGlySerAspGlyArgPheLeuArgGlyTyrLeuGlnTyrAlaTyrAspGlyArgAspTyrIleAlaLeuAsnGlu
CCCTCAGGCTTCACACAGATCCAGTGCATGATGGCTGTAAGTGGGGTCCGACGGGGCTTCCTCCCGGGTACCTGCAATACCCATACGACGGCCGGATACATGCTCCTGAACGAA- 840
AspLeuLysThrTrpThrAlaAlaAspValAlaAlaIleIleThrArgArgLysTrpGluGlnAlaGlyAlaAlaGluTyrTyrArgAlaTyrLeuGluAlaGluCysValGluTrpLeu
GACCTGAAACGTTGGAGCCGAGCGGAGCTGGCGGCATATCACCCAGCAGAGTGAGGAGCGGTGGTCTCGAGAGTATACAGGGCCCTACCTGGAGCCGAGTCGCTGGAGTGCTC- 960
LeuArgTyrLeuGluLeuGlyLysGluThrLeuLeuArgThrA
CTCAGTACTGGAGCTGGGAGGAGACGCTGCTGGCAGAGGTGACGGGCGCGGCGAGCTCCCTCCCTCTAGGCCTCGGCGGGGGTTCAGTCTGGGGAAGAAAGAAACCTCAGC-1080
TGGGTGTATGCCCX450BPXXXTGGCCCTTGGCACTCAGAAATCGCCCTGCCCTGCTGCTCGCCGCTGGGATTAAGGGCTGGGCCACCACCCTGGGCAGAAAGGTTCTCTGTG-1200
AGTTTAAATGTTTTCGGACAGGAATTAACATCCAGACTCCTTGATATCCCTGGTCCGACCCAGTACAGTGCCTTCCTGTGAAATCAGGACTGGACTCTGAGACAGAGGCTCCT-1320
CTGAATCCAGGGCTGAGTGAGAGGGGAAACACACACCTTGGAGCCCAGCTGTGTCAGTGAGTGCCTGGGTCACGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC-1440
TTCGCCCCAGAGTGGGGCTGGAGTCAATTTCTCTGGCTGAGTGTCAGAGGTTCCACACAATTTTCGTGACACATCCCTGATGGCTGTTCAGTGGAGTCAAGTTCGATCG-1560
GCAAGATGACCCAGAGTGGTTCAGTCAATGGTGACCTCTCCAGTAGCATAGCTGCTGTAATTTAGATAAGAACTCAAAACACAIATGAAATTCCTATTTTCATTCATCT-1680
TCCATATATAGTACTATCTCGTCTATGAACATCACATAGGATGACCATGTGACCCACTXX130BPXXXCTGACAGGTCACTAGTGCATGACAGTTGAAGTGTTCATACAGAC-1800
ACATAGTTCAGTGAATTAATTAATTTTGTCTTACGATTTTCAGTITGCTTTTATTTTGTGTTGTTGTTTATGTATGGAAGTACACTAGTGCATGAGTTCAGTTCAGTTC-1920
TGCCTGCTCGTGGTTGTGGGAAITGAAITTTTTCAGGGCCCTGCTGTGCTGATCGGCCCGCTCACTCCGCTCAACTCTAAGGCTCAACTGCTGGCTCAGCTCGC-2040
ATGTTGCTCGTGGTATTATTTATTTTACATAAATGACTACACTAGCTGACTTCGATGCAACCAGGAAGCGGTTCAGTCTCATACAGTCACTGTTAGGCCAGCTGGT-2160
GCTGGGCTTGAACCTCAGGACCTTCAAAGACGATTCAGTGCCTTATACCCTTAAAGCACTCTCCCGAGTCTCAGTTGCTTCTTAAATGTCGGATTCTTGAATCTCCACACAGAT-2280
ProProLysThrHisValThrHisHisProGlySerGluGlyAspValThrLeuArgCysTrpProLeuGlyPheTyrProAlaAspIleThrLeuThrTrpGlnLeuAsnGlyGluGlu
CCCCAAAGACACATGTGACCATCACCAGGATCTGAAGGTATGTACCCCTGAGGGTTGTCGGCCCTGGGCTTACCCTTGTGACATCACTCTGAGCTGAGCTGAGCTGAGGG-2400
LeuThrGlnAspMetGluLeuValGlnThrArgProAlaGlyAspGlyThrPheGlnLysTrpAlaSerValValProLeuGlyLysGluGlnAsnTyrThrCysHisValTyrHis
CTGACCCAGGACATGGAAGCTGGTGGAGACACAGGCCCTGACGGGATGGAACCTTCCAGAGTGGGACTTCAGTTCGCTGTTGGGAGGAGCAGAAATACACATGCTGTGATC-2520
GluGlyLeuProGluProLeuThrLeuArgTrpG
GAGGGCTGCTTGAAGCCCTCAGCTGAGATGGGTAAGGAGAGTGTGGTGCAGAGCTGTGGTCAGGAAAGCTGGAGCCTCTTACAGACCTTGAGTTGGTTCAGGCTGAGAGCTGGGG-2640
luProProSerThrAspSerIleMetSerHisIleAlaAspLeuLeuTrpProSerLeuLysLeuTrpTrpTyrLe
TCATGACCCCTCACCCTCTTTTCCTGTACCTTCCTTCGCCAGAACCTCGCTCCTTCACCTGACTCTATCAATGACACATGCTGATGCTGTTGGCCTCATTAAGCTTCGTTGGTATT-2760
utr
GTGATGAAGAGAAGGAGAAACACGGTGGGAAAGGGCAGGTCAGTTTCTCTCAGGCTCCTTTCAGGTGTCCTGCTCCTTCAATTAATCCAAACACAGCCACCCCAATGCTACTG-2880
ACTCAACATGGCTTCTCCTCAGTTCGGGAACCTTCCAGTGTCAAGATCTCTTTGAACCTCCTCAGTCTTTCCTTCCAGAGGTAAGAAAGAGGAGCATTGCTCCCGTCCAGGTTA-3000
GTGTCGGGACAGGATGTCCGCGGACATTGCAGTGAAGCTGGAGAGCTTGGGAGCTTCGGAAATCAATAACTTCCAGAGAAATCTTCCGCGCTTTCGGTTTTCAGTTCGATGA-3120
ATACATCTCTATACATTTCTTAACTAGGACGAGGATAGCTCCAGAGCTCTGATACCGCTTCGCAAGTTCGAAAGTTCGAAAGTTCAGGAGGAGAGCAGGTTGATACGA-3240
CTGGGTTTCAGGAGTCCAGAAATCCCTGTGATGTAGTCAATGGTTGTTGGGGTGGCTGACTTCATAGTAGTGGTTCATGACTGACTTCTGCTAGTTCAGAACACCTTACTGAGTGC-3360
ACTGAGTACAGCAGTGTTCAGTCTCTCTGTCAGATCCAGAGCCCTCAGTTCCTTTCAGCAACAGCTGCTGATGCTTCCTGTTTCTGTCGCTCAGTTCAGAACCTTCTGA-3480
CGCCAGCTGCCCTCCACACAGGACCCCTGCTGATTCGCTTCCTCTCAGACTGACCTTCCACCTCCAGCCAAACACTGAGGAGCAGCTGATCTGTCGTGAGCCTCACCTAC-3600
CTAAGCTGACGXXXXXXXCTCCACACTGAGAAATGAGAAITCGGATGTAACCTTATTTGTCATATCTTGAATTCAGGGCTTGAATGACAGGTAACATAAAGAAATAGAAATCTAGAG-3720
TTTCTGGAGGAAATAATGCGGATGGGAACTTCGGGAAASARXTAGRRATXGXATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT-3811

```

Figure 2. Nucleotide Sequence of the Q10 Gene  
Nucleotides are numbered from the 5' Bam HI site. Two gaps in the sequence are shown in the form XXX450BPXXX (i. e., a gap of approx. 450 bp). Small gaps, close to labeled restriction sites, are also depicted by XXX. As characters depicting gaps are also numbered in the figure, the numbers only give distance from the 5' end of the sequence up to the first gap (base number 1094). S = purine, R = pyrimidine. Predicted amino acid sequences for each of the five translated exons (see text) are shown. Untranslated exon 5, 6, and 7 sequences ending at the polyadenylation signal sequence (AATAAA) are underlined. The seven bases that differ in the Q10 and pH16 sequences are shown in bold, lower case characters. The differences are; base 291 T→C, base 406 G→A, base 527 C→T, bases 2784, 3371, 3742, 3748 are all G→A. Only two of these base changes cause amino acid coding changes; base 291, Met→Thr and base 527 His→Tyr.

view of the remarkably high sequence conservation and the identical, unusual structure of the Q10 gene and pH16 cDNA sequences, we believe that these two sequences are derived from allelic copies of a class I gene in C57BL/10 and SWR/J mice.

**Structure of the Q10 Gene**

Comparison of the Q10 sequence with other class I sequences defines the exon structure of the Q10 gene. Six exons (Figures 1a and 1b) are highly homologous to the pH16 (Kress et al., 1983a) and pH105 (unpublished) cDNA sequences. Exon 1 sequences, which encode the N-terminal signal peptide, are not present in the cDNA sequences, since they extend only up to amino acid codon 3 (amino acid 1 is defined as the N-terminal amino acid of the  $\alpha_1$ , or N, protein domain; Coligan et al., 1981). We detected, however, a 64 base sequence, starting 18 bases from the 5' end of the sequence in Figure 2, which is probably Q10 exon 1, based on the following criteria: it is bounded at the 5' end by an ATG codon and at the 3' end by a tetranucleotide sequence (AGGT) that fits the consensus sequences for a splice donor site (Mount, 1982); it is highly homologous to exon 1 sequences of other class I genes (Figure 1c); and it would encode 21

mostly hydrophobic amino acids.

Not all of these seven exon sequences (Figure 1a) encode amino acid sequences, since there is a TGA translational termination codon (base no. 2762) in the reading frame of exon 5 (predicted amino acid sequences are shown in Figure 2). Thus the 3' part of exon 5 and all of the exon 6 and 7 sequences are untranslated in mRNA products of the Q10 (Figure 1b, top line) and pH16 (Kress et al., 1983a) genes. This contrasts with the structure of other class I genes, such as the H-2K<sup>b</sup> (Weiss et al., 1983a) and H-2L<sup>d</sup> (Moore et al., 1982) genes, which encode classical transplantation antigens. The amino acid sequences of these antigens is encoded by eight exons (Figure 1b, lower line), seven of which are homologous and have similar lengths to the exons in the Q10 and pH16 cDNA sequences. As reported previously (Cosman et al., 1982a; Kress et al., 1983a), the missing exon sequence in pH16 corresponds to exon 7 in the prototype (i. e., H-2K<sup>b</sup> and H-2L<sup>d</sup>) H-2 class I gene (Figure 1b). Inspection of the Q10 sequence in the region between exons 6 and 8 (the 3' exon of pH16 is homologous to exon 8 of the prototype class I gene) reveals a 39 base sequence, which is homologous to prototype exon 7 sequences (Figure 3). Whereas this sequence (shown as 7' in Figure 1a) is

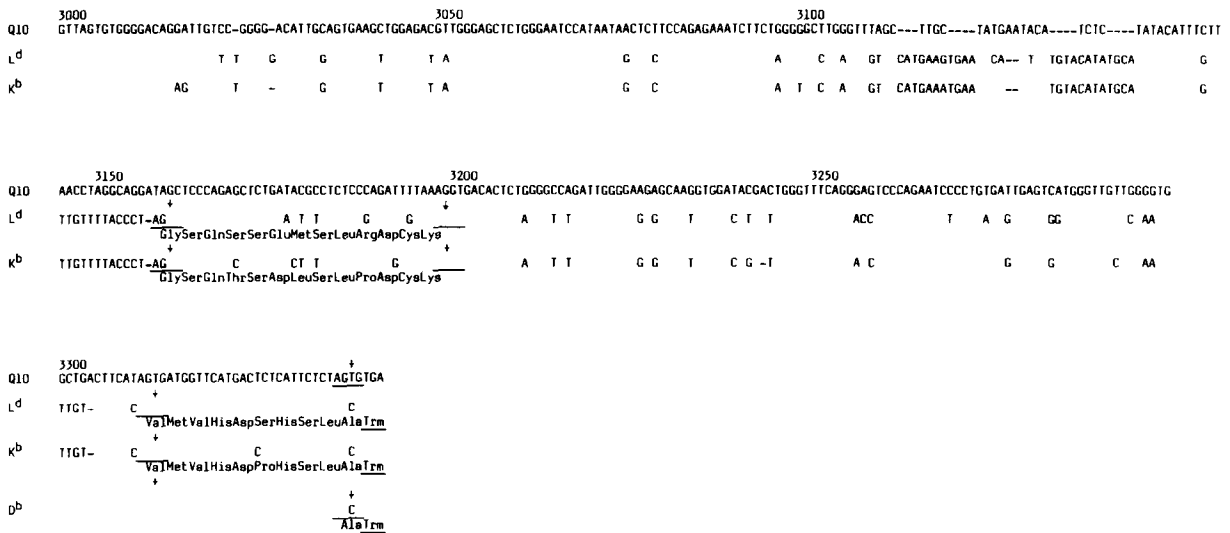


Figure 3. Sequence Comparisons at the 3' End of the Q10 Gene

Sequence of the Q10 gene between exons 6 and 8 compared to equivalent regions in the H-2L<sup>d</sup> (Moore et al., 1982) and H-2K<sup>b</sup> (Weiss et al., 1983a) genes. Nucleotide differences between each gene and the top line (Q10) only are shown. Splice acceptor and donor sites are underlined. Arrows indicate the first and last bases of an exon sequence. Only the 5 bases in exon 8 of the H-2D<sup>b</sup> (Reyes et al., 1983) are compared here.

bounded at the 3' end by a tetranucleotide (AGGT) that fits the consensus sequence for a splice donor site, it is bounded at the 5' end by a sequence (TAGC) that does not fit the consensus sequence for a splice acceptor site (Mount, 1982). In addition, sequences upstream of this site, in the region homologous to intron 6 of prototype class I genes, are organized differently in the Q10 gene (Figure 3). These differences, and/or the absence of a functional splice acceptor site, may explain why prototype exon 7 sequences are not present in mRNA transcripts and cDNA clones (pH16) derived from the Q10 gene. Similar findings have been reported previously for a putative class I pseudogene (Steinmetz et al., 1981) that was isolated from a Balb/c (H-2<sup>d</sup> haplotype) genomic library. The sequence comparisons in Figure 3 also show that the first base of Q10 exon 8 (base 3337) is 27 bases downstream of the first base of H-2K<sup>b</sup> exon 8 (equivalent to base 3310 in the Q10 sequence). In this respect, the Q10 gene is similar to the H-2D<sup>b</sup> gene (Kress et al., 1983a; Reyes et al., 1983). The reason for the use of the downstream splice acceptor site is not clear since a tetranucleotide sequence (AGTG) that fits the consensus sequence for a splice acceptor site is present at the same place in the Q10 gene as in the H-2K<sup>b</sup> gene. There is, however, a T instead of a C immediately 5' to this tetranucleotide in the Q10 gene, and further 5' is a stretch of 6 bases that differ in the Q10 and H-2K<sup>b</sup> genes.

### Q10 Locus Genes Are Much Less Polymorphic Than Genes at the H-2K Locus

The high conservation (99.5%) of the Q10 and pH16 sequences is unusual even for allelic class I gene sequences. For example, Weiss et al. (1983a) showed that H-2K<sup>b</sup> and H-2K<sup>d</sup> allelic sequences are much less conserved, mainly due to diversity in the 5' exons, which we

believe to have occurred through gene-conversion events. Clearly, the Q10 and pH16 sequences show no evidence of similar mechanisms acting on these alleles. The similarity, however, of the two sequences could be due to a close evolutionary relationship between the C57BL/10 and SWR/J mice such that there has been insufficient time to generate diversity of these alleles by gene-conversion events. To test this, we have compared the allelic H-2K sequences from the C57BL/10 and SWR/J mice. pH22 was isolated from the same SWR/J liver cDNA library as pH16 (Cosman et al., 1982a) and was sequenced as described in Experimental Procedures. The sequences of two exons (exon 5 and exon 8) revealed that this clone is derived from the H-2K<sup>a</sup> gene. Class I DNA sequences in these exons are specific for each class I gene (i. e., they are locus-specific sequences; Kress et al., 1983b). Comparison of the H-2K<sup>b</sup>, H-2K<sup>d</sup>, and pH22(K<sup>a</sup>) sequences (Table 1) reveals that the H-2K<sup>b</sup> and H-2K<sup>a</sup> alleles are more homologous, in all exons, than the H-2K<sup>b</sup> and H-2K<sup>d</sup> alleles—suggesting that C57BL/10 and SWR/J mice are more closely related than C57BL/10 and Balb/c mice. Since selective pressure is probably exerted least in the 3'UT region, the divergence in this region is probably the best available measure of the genetic drift between these two pairs of mouse strains (0.7% and 3.0%, for H-2K<sup>b</sup>/H-2K<sup>a</sup> and H-2K<sup>b</sup>/H-2K<sup>d</sup>, respectively). The low levels of H-2K<sup>b</sup>/H-2K<sup>a</sup> diversity, which might be attributable to genetic drift in exons 4, 6, 7, 8, and 3'UT (~0.6% over all these regions), is comparable with the low level of diversity between the Q10 and pH16 sequences (~0.5%), suggesting that divergence in both pairs of alleles in these exons is due to genetic drift. This figure contrasts with the higher diversity in exons 2, 3, and 5 of the H-2K alleles. The level of diversity (6.8%–9.4%) is similar to the diversity observed (5.8%–11.2%) between the H-2K<sup>b</sup> and H-2K<sup>d</sup>

Table 1. Comparison of H-2K Allelic Sequences from Three Different H2 Haplotypes

DOMAIN	EXON	LENGTH	TOTAL BASE CHANGES <sup>a</sup>		% TOTAL <sup>c</sup>	
			K <sup>a</sup> /K <sup>b</sup>	K <sup>d</sup> /K <sup>b</sup>	K <sup>a</sup> /K <sup>b</sup>	K <sup>d</sup> /K <sup>b</sup>
L	1	-	-	5(3)	-	7.8
$\alpha_1$	2	(213) <sup>b</sup>	18(7)	28(17)	8.4	10.3
$\alpha_2$	3	276	26(14)	31(22)	9.4	11.2
$\alpha_3$	4	276	2(0)	17(8)	0.7	6.1
TM	5	117	8(5)	9(5)	6.8	5.8
CY2	6	33	0	2	0	6.0
CY3	7	39	0	1	0	2.6
CY4	8	32	0	0	0	0
3'UT	(8)	423	3(0)	13(0)	0.7	3.0

(Footnotes)

- a. Number of clustered bases (Weiss et al., 1983a) is shown in brackets.
- b. This is the maximum extent of available pH22 (H-2K<sup>a</sup>) sequence for exon 2.
- c. Figures for the H-2K<sup>b</sup>/H-2K<sup>d</sup> comparison are taken from Weiss et al., 1983a.

alleles in these exons. In fact, when we correct these figures for the different sequence divergences observed in the 3'UT regions of the two pairs of alleles (3.0% - 0.7% = 2.3%) the figures are comparable. Thus exons 2, 3, and 5 in all three H-2K alleles have been equally exposed to mechanisms that generate high diversity. This implies that since similar polymorphism is not observed in the Q10 and pH16 sequences, the Q10 gene and its alleles are either not exposed to similar mechanisms, or selective pressures prevent the fixation of mutated alleles at this locus.

## Discussion

The Q10 gene is unlike prototype H-2 class I genes such as H-2L<sup>d</sup> and H-2K<sup>b</sup>, which encode classical transplantation antigens, because it has an unusual exon 5 sequence that would cause premature termination of translation in the transmembrane domain of a Q10 polypeptide product. This phenomenon has been noted previously for a cDNA that was isolated from a cDNA library made from SWR/J (H-2<sup>a</sup> haplotype) liver RNA. When the Q10 gene sequence is compared to this cDNA, the two sequences are found to be almost identical (99.5% conserved). This is a remarkable finding for class I sequences obtained from mice carrying different H-2 haplotypes. We propose, on the basis of this high homology and the unusual exon 5 sequence, that the cDNA sequence is derived from a gene that is allelic to the Q10 gene in SWR/J mice. In this case, the Q10 gene is also unusual in having a different pattern of splicing at its 3' end, removing sequences that are homologous to exon 7 of prototype class I genes from any mRNA product. The absence of a functional splice acceptor site preceding the 39 base Q10 sequence, which is homologous to prototype exon 7 sequences, may explain why these sequences are not found in mRNA transcripts of the Q10 gene. A similar finding has been reported previously for a putative pseudogene (Steinmetz et al.,

1981) that maps to the Qa2,3 region of Balb/c (H-2<sup>d</sup> haplotype) mice.

Because of the unusual exon 5 sequence, Kress et al. (1983a) have suggested that the gene from which the pH16 cDNA is derived encodes a secreted class I antigen. By implication, this may also be true for the Q10 gene described here. In addition, Cosman et al. (1982b) have shown that mRNA sequences homologous to the pH16 cDNA sequence are present at high levels in liver cells of most mice strains (including C57BL), but are absent from all other tissues tested. Thus the Q10 gene, which maps to the Qa2,3 region, may encode a novel, liver-specific, class I antigen of unknown function that has not been detected immunogenetically. We have evidence, from polymorphic restriction enzyme mapping studies, that the Q10 gene is not polymorphic in different inbred mouse strains. Indeed, we have noticed that the restriction map in and around a class I gene isolated from a Balb/c (H-2<sup>d</sup> haplotype) mouse library (cluster 9 in Steinmetz et al., 1982) is almost identical with the restriction map in and around the Q10 gene. This gene has also been mapped to the Qa2,3 locus (Winoto et al., 1983). Thus lack of polymorphism at the Q10 locus may explain why the product has not been detected serologically. Perhaps more of the class I genes isolated by molecular cloning encode class I polypeptides that have, so far, been undetected by immunogenetic analysis. The data of Goodenow et al. (1982) suggests that this may be the case for several class I genes in the Balb/c mouse.

Three explanations could account for the lack of high allelic polymorphism, which is characteristic of H-2 class I genes, in the genes at the Q10 locus. First, C57BL/10 and SWR/J mice may be closely related such that the mechanisms that generate high allelic polymorphism have had insufficient time to generate high polymorphism in alleles from these mice. Second, the Q10 gene may not be a recipient of intergenic exchange of genetic material between class I genes; a process that we (Weiss et al., 1983b) and others (Pease et al., 1983) believe accounts for much of the high allelic polymorphism at the H-2K locus. Third, the genes at the Q10 locus may be recipients of exchanged genetic material but, due to selective pressure, mutated Q10 alleles are not fixed in the genome of the mouse. To eliminate the first possibility, we have also made comparisons between the allelic sequences H-2K<sup>b</sup> and H-2K<sup>a</sup>. This suggests that the C57BL/10 mice may be more closely related to SWR/J than to Balb/c mice. It is striking, however, that the H-2K<sup>b</sup> and H-2K<sup>a</sup> sequences are almost as polymorphic as the H-2K<sup>b</sup> and H-2K<sup>d</sup> sequences in exons 2, 3, and 5, in sharp contrast to our findings from the Q10 and pH16 sequence comparisons. This shows that the first explanation to account for lack of allelic polymorphism at the Q10 locus is not correct and that it must be due to fundamental differences in the evolution of class I gene sequences at this locus, compared to genes at the H-2K locus. This conclusion is particularly intriguing since we have discovered that the Q10 gene is a potential donor gene for gene-conversion events at the

H-2K locus of the C57BL mouse (Mellor et al., 1983). Thus, either these events involve a nonreciprocal intergenic exchange of genetic material, or reciprocal exchange does take place but mutant Q10 alleles created by this mechanism cannot be fixed in mice populations. Clearly, we cannot discriminate between these possibilities until we are able to analyze the donor gene from an intergenic transfer event immediately after the event has taken place. It is, however, intriguing to speculate whether the Q10 gene encodes a liver-specific, secreted class I antigen that has a critical, but as yet undefined, function and whether, if lost, this puts the mouse at a selective disadvantage.

### Experimental Procedures

#### General Methods

Preparation of cosmid libraries (Grosveld et al., 1981) and isolation of cosmids containing class I genes (Mellor et al., 1982) from these libraries have been described previously.

#### DNA Sequence Analysis

DNA sequence analysis of the three Bam HI fragments containing parts of the Q10 gene subcloned in the vector pBR327 was carried out as described by Maxam and Gilbert (1980). End-labeled fragments were generated by restricting DNA end-labeling with T4 polynucleotide kinase in the presence of  $\gamma$ - $^{32}\text{P}$ -ATP following dephosphorylation of the 5' overhang (Bam HI and Bgl II) with phosphatase (Maxam and Gilbert, 1980) or terminal transferase in the presence of  $^{32}\text{P}$ -cordycepin (Tu et al., 1980; kits obtained from Amersham International and New England Nuclear) for 3' overhangs (Kpn I, Sac II, and Pst I). End-labeled fragments were spun through a Sephadex G50 column and restricted with a second enzyme to generate single end-labeled fragments. Labeled fragments were separated on low-melting agarose gels and extracted using phenol. Chemically modified fragments were separated on 3x 50 (or 60) cm long acrylamide gels (6%) run for different times such that 250–300 bases from the labeled end could be read routinely.

Sequence analysis of the cDNA clones pH105 and pH22 was carried out using the primer-extension method described previously (Kress et al., 1983a).

#### Acknowledgments

We thank Richard Tizard (Biogen) for helpful advice on DNA sequencing, James Devlin (Biogen) for critical reading of the manuscript, and Madlyn Nathanson (Biogen) for preparing the manuscript.

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

Received November 7, 1983; revised November 15, 1983

#### References

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

Cosman, D., Khoury, G., and Jay, G. (1982a). Three classes of mouse H-2 messenger RNA distinguished by analysis of cDNA clones. *Nature* 295, 73–76.

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

Goodenow, R. S., McMillan, M., Nicolson, M., Taylor-Sher, B., Eable, K., Davidson, N., and Hood, L. (1982). Identification of the class I genes of the mouse major histocompatibility complex by DNA-mediated gene transfer. *Nature* 300, 231–237.

Grosveld, F. G., Dahl, H. H. M., deBoer, E., and Flavell, R. A. (1981). Isolation of  $\beta$ -globin related genes from a human cosmid library. *Gene* 13, 227–237.

Klein, J. (1975). *Biology of the Mouse Histocompatibility-2 Complex*. (New York: Springer-Verlag), pp. 192–230.

Kress, M., Cosman, D., Khoury, G., and Jay, G. (1983a). Secretion of a transplantation-related antigen. *Cell* 34, 189–196.

Kress, M., Liu, W.-Y., Jay, E., Khoury, G., and Jay, G. (1983b). Comparison of class I (H-2) gene sequences: derivation of unique probes for markers of this multigene family. *J. Biol. Chem.*, in press.

Maxam, A., and Gilbert, W. (1980). DNA sequencing procedures. *Meth. Enzymol.* 65, 455–560.

Mellor, A. L., Golden, L., Weiss, E., Bullman, H., Hurst, J., Simpson, E., James, R., Townsend, A. R. M., Taylor, P. M., Schmidt, W., Fertuga, J., Leben, L., Santamaria, M., Atfield, G., Festenstein, H., and Flavell, R. A. (1982). Expression of murine H-2K<sup>b</sup> histocompatibility antigen in cells transformed with cloned H-2 genes. *Nature* 298, 529–534.

Mellor, A. L., Weiss, E. H., Ramachandran, K., and Flavell, R. A. (1983). A potential donor gene for the bmi gene conversion event in the C57BL mouse. *Nature*, in press.

Michaelson, J., Boyse, E. A., Chorney, M., Flaherty, L., Fleisner, I., Hammerling, U., Reinisch, C., Rosenson, R., and Shen, F.-W. (1983). The biochemical genetics of the Qa-Tla region. *Transplantation Proc.*, in press.

Moore, K. W., Taylor-Sher, B., Sun, H.-T., Eakle, K. A., and Hood, L. (1982). DNA sequence of a gene encoding a Balb/c mouse L<sup>d</sup> transplanted antigen. *Science* 215, 679–682.

Mount, S. M. (1982). A catalogue of splice junction sequences. *Nucl. Acids Res.* 10, 459–472.

Pease, L. R., Schulze, D. H., Pfaffenbach, G. M., and Nathanson, S. G. (1983). Spontaneous H-2 mutants provide evidence that a copy mechanism analogous to gene conversion generates polymorphism in the major histocompatibility complex. *Proc. Nat. Acad. Sci. USA*, 80, 242–246.

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

Reyes, A. A., Schold, M., and Wallace, R. B. (1983). The complete amino acid sequence of the murine transplantation antigen H-2D<sup>b</sup> as deduced by molecular cloning. *Immunogenetics* 16, 1–9.

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

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

Tu, C.-P. D., and Cohen, S. N. (1980). 3'-end labelling with ( $\alpha$ - $^{32}\text{P}$ ) cordycepin-5'-triphosphate. *Gene* 10, 177–183.

Weiss, E. H., Golden, L., Zakut, R., Mellor, A. L., Fahrner, K., Kvist, S., and Flavell, R. A. (1983a). The DNA sequence of the H-2K<sup>b</sup> gene: evidence for gene conversion as a mechanism for the generation of polymorphism in histocompatibility antigens. *EMBO J.* 2, 453–462.

Weiss, E. H., Mellor, A. L., Golden, L., Fahrner, K., Simpson, E., Hurst, J., and Flavell, R. A. (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* 307, 671–674.

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