

Identification of a Recombinogenic Major Histocompatibility Complex Q Gene with Diverse Alleles

By Mary K. Cullen,*[‡] Lynne A. Lapierre,[‡] Krishna V. Kesari,*[‡] and Jan Geliebter*[‡]

From the *Howard Hughes Medical Institute; and [‡]The Rockefeller University, New York, New York 10021

Summary

Structural diversity enables class Ia molecules to present a diverse repertoire of peptides to the T cell receptor. This diversity is thought to be generated by recombinations between class I genes. We have found that two class Ib Q2 alleles exhibit extremely high sequence diversity, even higher than class Ia alleles. Clustered nucleotide differences between Q2^b and Q2^k suggest that this sequence diversity was generated by microrecombinations between Q2 genes and other class I genes. The relatively high expression of Q2^b in the thymus may be significant and perhaps suggests a novel role for a Q2^b product in the education and selection of the T cell repertoire.

Extensive sequence diversity among alleles of the polymorphic MHC class Ia loci (H-2 K, D, L; HLA-A, B, C) provides the structural basis for the presentation of a wide array of peptide antigens to the cellular immune system (1). Despite structural homology to class Ia, the restricted tissue distribution, limited polymorphism, and sequence conservation among class Ib loci (Q, TL, M) were thought to preclude an immunologic role for these molecules. More recently, M molecules have been found to present prokaryotic and mitochondrial peptides to the immune system (2, 3). Further, TL products have been identified as target molecules for both α/β and γ/δ T cell clones (4–6). Although known to be targets of H-2-unrestricted CTL, Q molecules have an as yet unknown function (7). We now report that alleles of the Q2 locus exhibit greater sequence diversity than either class Ia or class Ib alleles and appear to have participated in many gene conversion-like (microrecombination) events. Q2^b expression in the thymus suggests a novel role for this unusual class Ib in the formation of the T cell repertoire.

Materials and Methods

Genes. The Q1^b and Q2^b genes from cosmid clones LSHT36 and H26, respectively, were obtained from Dr. R. Flavell (Yale University, New Haven, CT) (8), and were digested with BamHI, subcloned into pBLUESCRIPT (Stratagene, La Jolla, CA), and dideoxy sequenced (Sequenase; United States Biochem. Corp., Cleveland, OH) using gene- and vector-specific primers.

RNA Preparation and PCR Analysis. The preparation of whole cellular RNA was as previously described (9). cDNA was prepared by incubating 2 μ g of RNA with 100 ng of Q2^b 3' primer (CCTGCGCTTCTCTGCAAG, positions 152–157) or K^b 3' primer

(tacgagaatCTGAGTCTCTCTGCTCACCAGC, positions 150–157) in 20 μ l of 1 \times PCR buffer (PEXPRESS, Norwalk, CT) with additional MgCl₂ to 6.5 mM and 20 U reverse transcriptase (RT) (Life Sciences, Inc., St. Petersburg, FL) for 30 min at 50°C, followed by 95°C for 5 min. Nucleotide sequence in lower-case letters at 5' of the oligonucleotides represents added restriction enzyme sites. After the addition of 400 ng 3' primers, 500 ng of Q2^b 5' primer (tctggagaattCCCTGAAACTGACTGAGACA, positions –3 to –9) and K^b 5' primer (ttcaagaattCCCTGGCTCCGACTCAGA, positions –3 to –9) 2 U Taq polymerase (PEXPRESS), and 1 \times PCR buffer to 100 μ l, amplification proceeded by incubation at 94°C for 30 s, 55°C for 30 s, and 72°C for 1 min. After 15 cycles, 5 μ l was removed and used as input DNA for an additional 25 cycles. Fractionation by agarose gel electrophoresis, transfer to GenScreen (New England Nuclear, Boston, MA), and hybridization analysis with specific probes was as previously described (9). Q2^b- and K^b-specific RT-PCR were carried out in the same or in different amplification tubes with similar results.

Results and Discussion

Q1 Alleles Are Conserved. Sequence analysis of the Q genes of the C3H mouse indicated that Q1^k, while potentially coding for an intact class Ib molecule, is one of the most divergent Q genes (10). Sequence analysis of exons 1–3 of the Q1^b gene demonstrates that the two alleles are almost identical, with only two nucleotide differences (Fig. 1 and Table 1). This finding extends the previous observation of allelic sequence conservation of Q loci (4, 11).

Q2^b and Q2^k Are Alleles and Exhibit Sequence Diversity. In contrast to Q1, a nucleotide sequence comparison between Q2^b and Q2^k reveals that the two alleles differ from each

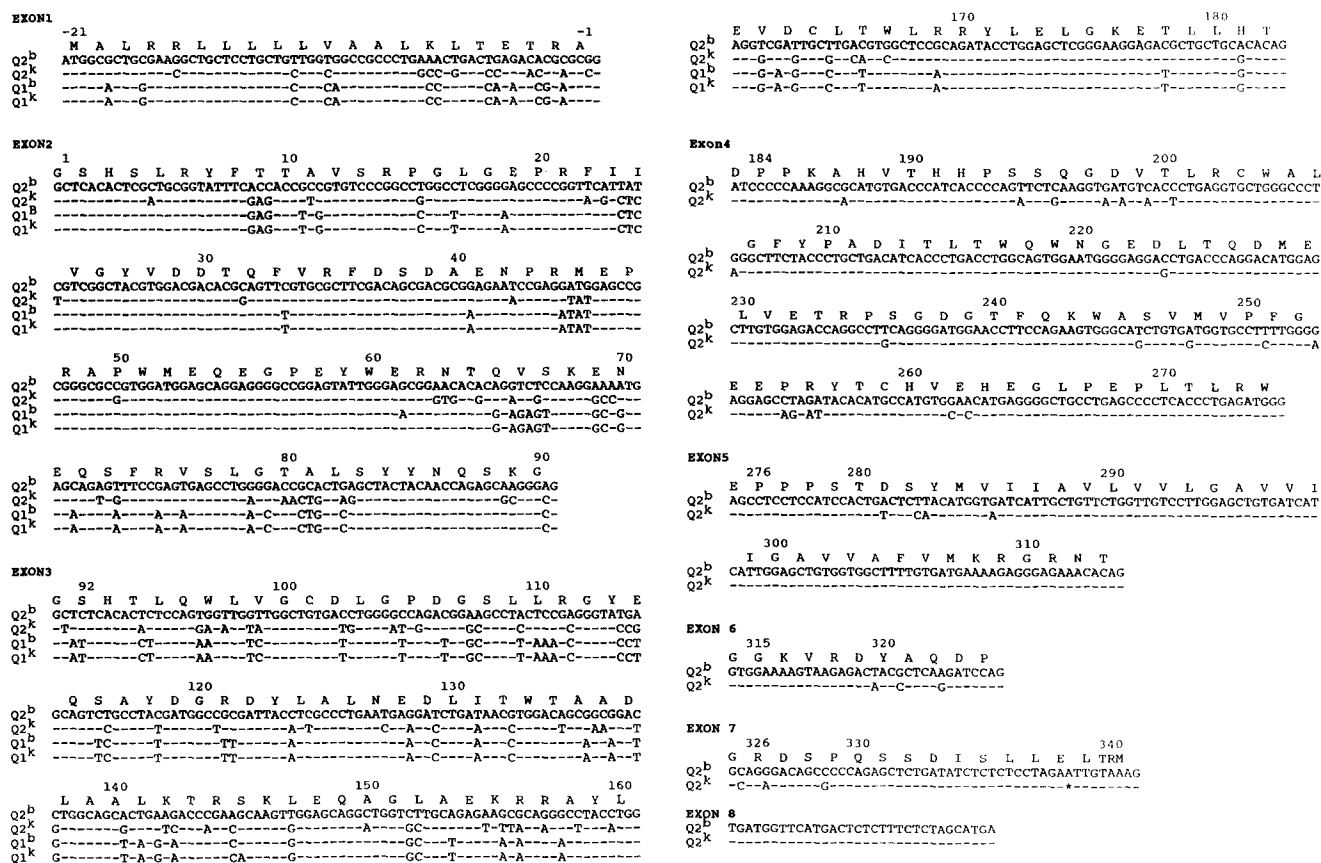


Figure 1. Sequence comparisons of Q1 and Q2 genes. Sequences are presented according to the exon organization of the K^b gene. Letters on top of nucleotide sequences are the symbols for amino acids coded for by Q2^b, and numbers on top of the symbols indicate residue number according to the K^b sequence (24). Negative numbers on top of exon 1 reflect a cleaved leader peptide. Q1^k and Q2^k sequences are from Watts et al. (10). Dashes beneath the Q2^b sequence indicate identity to Q2^b. The asterisk at position 338 of Q2^k indicates the position of a single base insertion in Q2^b. These sequence data are available from EMBL/GenBank data libraries under accession numbers X70381-X70390.

Table 1. Sequence Differences between Allelic Genes in the H-2^b and H-2^k Haplotypes

Alleles	No. of nucleotide differences in exons 2 and 3*	No. of amino acid differences in:		
		Domains 1 and 2 [†]	ARS [§]	Framework
K ^b /K ^k	51	27	14	13
D ^b /D ^k	63	31	14	17
Q4 ^b /Q4 ^k	1	0	0	0
Q10 ^b /Q10 ^k	1	1	0	0
Q1 ^b /Q1 ^k	2	1	1	0
Q2 ^b /Q2 ^k	97	46	27	19

DNA sequences other than Q1^b and Q2^b are from Kuhner and Goodenow (13). Amino acid sequences other than Q1^b and Q2^b are from Watts et al. (18).

* The number of differences between listed alleles in the 546 nucleotides that make up exons 2 and 3.

† The number of differences between listed alleles in the 182 amino acids that make up domains 1 and 2.

§ The number of differences between listed alleles in the 57 amino acids that make up the ARS in domains 1 and 2 (1).

|| The number of differences between listed alleles in the 125 amino acids that make up the non-ARS positions in domains 1 and 2.

other by 97 of the 546 nucleotides in exons 2 and 3 (comprising the $\alpha 1$ and $\alpha 2$ domains), resulting in 46 amino acid replacements (Fig. 1 and Table 1). This level of allelic diversity is even greater than that observed at H-2 loci (Table 1).

The allelic relationship between $Q2^b$ and $Q2^k$ is supported by both physical mapping studies and critical sequence homologies. Both genes have been cloned in cosmids that are part of large, defined contiguous genomic segments that encompass the entire Q region of their respective haplotypes (8, 10). $Q2^b$ and $Q2^k$ map as the first genes telomeric (~20 kb) to the Q1 genes in each haplotype (8, 10). Given the conserved nature of Q1 alleles as a benchmark, the mapping data support the idea that $Q2^b$ and $Q2^k$ are alleles.

Striking sequence homologies further support the allelic relationship between $Q2^b$ and $Q2^k$. Flanking $Q2^k$ (and $Q1^k$) are a unique 5' sequence of ~650 nucleotides (beginning 82 nucleotides upstream of the initiation codon and continuing in the 5' direction) and a unique 3' sequence of ~500 nucleotides (beginning ~100 nucleotides downstream of the termination codon) (10). Neither of these sequences has a counterpart in the GenBank data library (10). $Q2^b$ contains the almost identical unique sequences at precisely the same positions, suggesting an allelic relationship between the two genes (Lapierre, L. A., and J. Geliebter, manuscript in preparation).

Further, coding and intron sequence homologies also suggest an allelic relationship between $Q2^b$ and $Q2^k$. It has previously been shown that class I alleles exhibit sequence conservation in the 3' portion of their genes, while nonallelic genes often display sequence diversity (12, 13). For example, in exon 5 of the H-2^b and H-2^k haplotypes, K, D, Q4, and Q10 loci display >95% allelic sequence conservation, but <86% sequence identity between nonallelic genes (13). $Q2^b$ and $Q2^k$ are ~95% identical in exons 5-8, and introns 5-7, which is in line with expectations for alleles (Fig. 1, and unpublished data). Therefore, based on sequence conservation and physical mapping it can be inferred that $Q2^b$ and $Q2^k$ are alleles.

With the exception of CCAAC (instead of CCAAT) and TATAA sequences, none of the usual H-2 and Q regulatory sequences have been found in $Q2^b$ or $Q2^k$, a situation also found in TL region genes (10, 11, and Lapierre, L. A., and J. Geliebter, manuscript in preparation). Splice donor and acceptor sites are intact in all seven introns (data not shown), and $Q2^b$ would be expected to code for a 339-amino acid integral membrane glycoprotein. It is also possible that $Q2^b$ could be anchored to the membrane by a phosphatidylinositol (PI) linkage as with the Qa-2 protein (gene Q7) (11). Due to a single base insertion in exon 7, $Q2^b$ translation would terminate at the TAA after position 339. $Q2^b$ has only one N-linked glycosylation site at position 86, having lost a site at position 256 that is common to most D and Q region genes, but not present in most K alleles (13).

Despite the extreme diversity of $Q2^b$, there are no obvious amino acid substitutions that would preclude its assuming the usual class I conformation. Most amino acid replacements are either conservative and/or found in other class I genes of various species. For example, Gly 79, Trp 219, and Arg

256 are not found in any other murine H-2, Q, TL, or M genes, but can be found in some or all HLA class Ia or Ib genes (11, 13-15). Other residues, such as Leu-124, were previously found only in M region molecules (13, 16). Conserved amino acid positions that are presumed to be critical in the folding of class I molecules and binding β_2 -microglobulin are also found in $Q2^b$. These include His 3, Thr 10, Pro 15, Asp 29, Gln 96, Cys 101, Gly 120, Asp 122, Cys 164, Pro 185, Phe 208, Tyr 209, Pro 210, Pro 235, Arg 234, Pro 235, Asp 238, Gln 242, Cys 259, Val 261, Leu 266, and Pro 269 (17-19). Taken together, these findings suggest

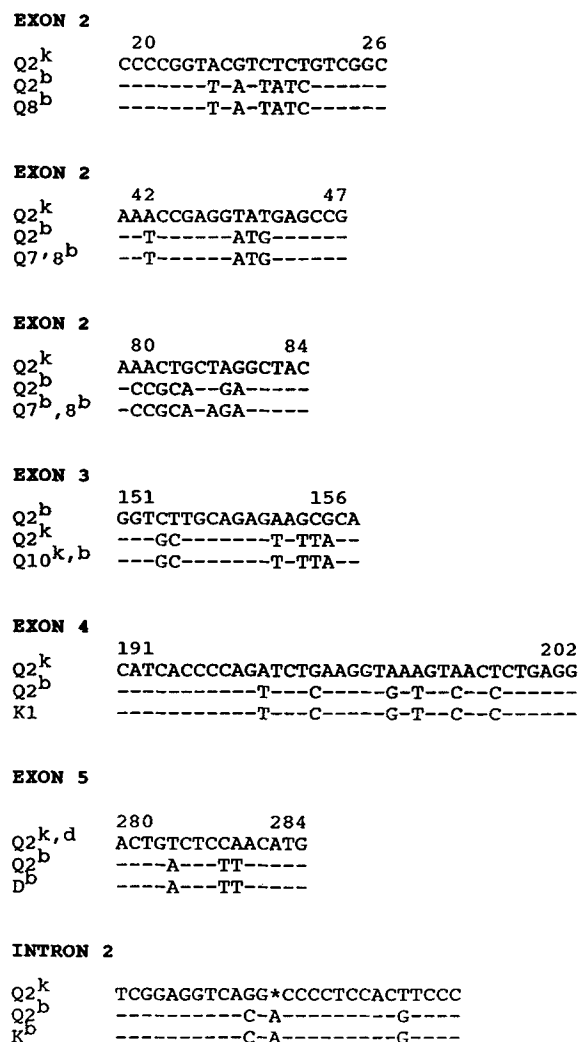


Figure 2. Potential microrecombination sites in the Q2 gene. Numbers on top of nucleotide sequences indicate residue number according to the K^b sequence. Dashes beneath the top sequence indicate identity to that sequence. Sequences other than $Q2^b$ are from Kuhner and Goodenow (13) and Brorson et al. (25). The asterisk in the intron 2 sequence of $Q2^k$ indicates the position of a single base insertion in $Q2^b$. Data are not meant to imply that all or any of the differences between $Q2^b$ and $Q2^k$ were generated since the establishment of inbred mouse strains, rather that potential donor sequences, are available.

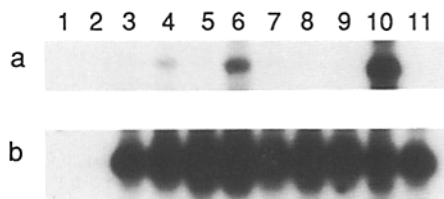


Figure 3. Transcription of the Q2^b gene. (a) Hybridization of RT-PCR products with a Q2^b-specific probe (TGGGAGACCTGTGTGTTCCG, positions 62–68) at 60°C identifies an ~550-nucleotide-long band. Similar results were obtained with another Q2^b-specific probe (GCCAACCAACCACTGGAGA, positions 94–100). Lane 1, no RNA; lane 2, brain; lane 3, heart; lane 4, kidney; lane 5, liver; lane 6, lung; lane 7, lymph node; lane 8, skin; lane 9, spleen; lane 10, thymus; lane 11, tongue. (b) Hybridization of RT-PCR products with a K^b-specific probe (GGTACATGGAAGTCGGCTAC, positions 21–27) at 60°C identifies an ~550-nucleotide-long band. This controls for the integrity and quantity of RNA in each sample. Samples in lanes are the same as in a.

that Q2^b may be expressed on the cell surface and function as a recognition structure.

Q2 Appears to Have Undergone Multiple Microrecombinations. Many of the nucleotide differences between Q2^b and Q2^k are clustered and can be found at homologous positions in other class I genes (Figs. 1 and 2). These characteristics are taken to reflect past microrecombination or gene conversion-like events with nonallelic class I genes (20). Kuhner et al. (21) analyzed sequences from 24 class I genes using a statistical algorithm and detected 25 past conversion events among all the genes analyzed. Visual examination of Q2^b, Q2^k, and six class I sequences of the H-2^b haplotype revealed evidence for more than six events (Fig. 2).

While most examples of past Q2^b microrecombination events are in exons 2 and 3, examples are seen in introns and the transmembrane exon 5 (Fig. 2). It is not always certain which Q2 gene was the recipient of genetic information. However, in exon 5, Q2^b differs from the published sequences of both Q2^k and Q2^d by the same nucleotides that are found in the D^b gene. This implies that Q2^b was probably the recipient of genetic information in a microrecombination event with a D^b-like gene.

Q2 Sequence Diversity Appears to Be the Result of Selection. The reassortment of class I sequences into H-2 genes has been suggested to provide the raw material for diversification, to be acted upon by natural selection (22). Selection for many allelic variants, each binding and presenting different repertoires of antigenic peptides, would result in the extant diverse and polymorphic class Ia alleles.

The lack of diversity and polymorphism of Q loci may reflect a unidirectional bias in conversion from Q genes to

H-2 genes (22). Alternatively, since most conversion events are lost in the absence of selection for the novel products, Q gene sequence conservation may simply reflect the lack of strong selection pressure for diversification (20). If the latter is correct, Q2 gene diversity and apparent maintenance of microrecombination products may reflect selection and imply a function more closely related to class Ia molecules than class Ib molecules. This notion is supported by two related observations; a very high percentage of amino acid replacements between the two alleles are in the antigen recognition site (ARS), and an unusually high percentage of nucleotide substitutions in the ARS result in amino acid replacements (see below).

The $\alpha 1$ and $\alpha 2$ domains of class Ia molecules contain 182 amino acids. Of these, 125 are in framework positions, and 57 are designated ARS positions (1). Of the ~30 amino acid differences between K^b and K^k or D^b and D^k, 50% are in the ARS, despite ARS positions constituting only 31% of positions (Table 1). Likewise, of the 49 amino acid differences between Q2^b and Q2^k in the $\alpha 1$ and $\alpha 2$ domains, 57% are in the ARS (Table 1). Thus, allelic differences between H-2 and Q2 molecules are predominantly in the ARS, where they affect peptide binding and interactions with the TCR.

A high ratio of nonsynonymous to synonymous nucleotide substitutions in the ARS is consistent with positive selection for sequence diversity. For class Ia, in the ARS, the rate of nonsynonymous substitutions per nonsynonymous site (d_N) is 21.4 ± 2.4 , while the rate of synonymous substitutions per synonymous site (d_S) is 13.8 ± 3.2 , indicating positive selection for diversification (23). Comparing Q2^b and Q2^k under the same analysis yields a d_N of 38.2 ± 6.8 and a d_S of 22.15 ± 8.2 , also indicating selection for diversity between the two alleles (Lorraine Flaherty, personal communication).

These data indicate that like class Ia, structural differences between Q2 alleles could result in functional diversity, and that these differences may be maintained by natural selection.

Q2^b Is Transcribed in Various Tissues. cDNA synthesis followed by RT-PCR of Q2^b mRNA indicates that Q2^b transcription is highest in the thymus (Fig. 3). Lower expression was also observed in the lung and kidney, and with prolonged exposure of the hybridized filter to film, most tissues showed a band of the appropriate size (PCR controls remained negative). Since RT-PCR is not quantitative and tissues were not perfused, weak expression in some tissues may reflect a disproportionate contribution of certain cell types. Yet, the above data do clearly show that Q2^b is transcribed and may be translated. The relatively high expression of Q2^b in the thymus may be significant and perhaps suggests a novel role for a Q2^b product in the education of and selection of the T cell repertoire.

We thank Dr. Lorraine Flaherty for the analysis of synonymous and nonsynonymous nucleotide substitutions in the Q2^b and Q2^k genes. We also thank Dr. Richard A. Flavell for generously providing cosmid clones of Q1^b and Q2^b.

This work was supported by the Howard Hughes Medical Institute.

Address correspondence to Jan Geliebter, Box 207, The Rockefeller University, 1230 York Avenue, New York, NY 10021. Mary Cullen's present address is the Division of Dermatology, Box 8123, Washington University School of Medicine, 660 South Euclid Avenue, St. Louis, MO 63110.

Received for publication 7 January 1993 and in revised form 17 March 1993.

References

1. Bjorkman, P.J., M.A. Saper, B. Samraoui, W.S. Bennet, 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.
2. Loveland, B., C. Wang, H. Yonekawa, E. Hermel, and K. Fischer-Lindahl. 1990. Maternally transmitted histocompatibility antigen of mice: a hydrophobic peptide of a mitochondrially encoded protein. *Cell* 60:971.
3. Pamer, E.G., C.R. Wang, L. Flaherty, K. Fischer-Lindahl, and M.J. Bevan. 1992. H-2M3 presents a *Listeria monocytogenes* peptide to cytotoxic T lymphocytes. *Cell* 70:215.
4. Ito, K., L. Van Kaer, M. Bonneville, S. Hsu, D.B. Murphy, and S. Tonegawa. 1990. Recognition of the product of a novel MHC TL region gene (27b) by a mouse gamma delta T cell receptor. *Cell* 62:549.
5. Milligan, G.N., L. Flaherty, V.L. Braciale, and T.J. Braciale. 1991. Nonconventional (TL-encoded) major histocompatibility complex molecules present processed viral antigen to cytotoxic T lymphocytes. *J. Exp. Med.* 174:133.
6. Vidovic, D., M. Roglic, K. McKune, S. Guerder, C. MacKay, and Z. Dembic. 1989. Qa-1 restricted recognition of foreign antigen by a gamma delta T cell hybridoma. *Nature (Lond.)* 340:646.
7. Forman, J. 1979. H-2 unrestricted cytotoxic T cell activity against antigens controlled by genes in the Qa/TLa region. *J. Immunol.* 123:2451.
8. 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.
9. Geliebter, J., R.A. Zeff, D.H. Schulze, L.R. Pease, E.H. Weiss, A.L. Mellor, R.A. Flavell, and S.G. Nathenson. 1986. Interaction between K^b and Q4 gene sequences generates the K^{bm6} mutation. *Mol. Cell. Biol.* 6:645.
10. 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.
11. Flaherty, L., E. Elliott, J.A. Tine, A.C. Walsh, and J.B. Waters. 1990. Immunogenetics of the Q and TL regions of the mouse. *Crit. Rev. Immunol.* 10:131.
12. Arnold, B.H. Burgert, A.L. Archibald, and S. Kvist. 1984. Complete nucleotide sequence of the murine H-2K^k gene. Comparison of three H-2K alleles. *Nucleic Acids Res.* 12:9473.
13. Kuhner, M.K., and R.S. Goodenow. 1989. DNA sequences of mouse and Qa genes. *Immunogenetics.* 30:458.
14. Parham, P., C.E. Lomen, D.A. Lawlor, J.P. Ways, N. Holmes, H.L. Choppin, R.D. Slater, A.M. Wan, and P.D. Ennis. 1988. Nature of polymorphism in HLA-A, -B, and -C molecules. *Proc. Natl. Acad. Sci. USA.* 85:4005.
15. Pohla, H., W. Kuon, P. Tabaczewski, C. Doerner, and E.H. Weiss. 1989. Allelic variation in HLA-B and HLA-C sequences and the evolution of the HLA-B alleles. *Immunogenetics.* 29:297.
16. Brorson, K.A., S. Richards, S.W. Hunt III, H. Cheroutre, K. Fischer-Lindahl, and L. Hood. 1989. Analysis of a new class I gene mapping to the Hmt region of the mouse. *Immunogenetics.* 30:273.
17. Fremont, D.H., M. Matsumura, E.A. Stura, P.A. Peterson, and I.A. Wilson. 1992. Crystal structures of two viral peptides in complex with murine MHC class I H-2Kb. *Science (Wash. DC)* 257:919.
18. Watts, S., C. Wheeler, R. Morse, and R.S. Goodenow. 1989. Amino acid comparison of the class I antigens of the mouse major histocompatibility complex. *Immunogenetics.* 30:390.
19. Grossberger, D., and P. Parham. 1992. Reptilian class I major histocompatibility complex genes reveal conserved elements in class I structure. *Immunogenetics.* 36:166.
20. Geliebter, J., and S.G. Nathenson. 1987. Recombination and the concerted evolution of the murine MHC. *Trends. Genet.* 3:107.
21. Kuhner, M.K., S. Watts, W. Klitz, G. Thompson, and R.S. Goodenow. 1990. Gene conversion in the evolution of both the H-2 and Qa class I genes of the murine major histocompatibility complex. *Genetics.* 126:1115.
22. Bregegere, F. 1983. A directional process of gene conversion is expected to yield dynamic polymorphism associated with stability of alternative alleles in class I histocompatibility antigens gene family. *Biochimie (Paris).* 65:229.
23. Hughes, A.L., and M. Nei. 1989. Evolution of the major histocompatibility complex: independent origin of nonclassical class I genes in different groups of mammals. *Mol. Biol. Evol.* 6:559.
24. Nathenson, S.G., H. Uehara, B.M. Ewenstein, T.J. Kindt, and J.E. Coligan. 1981. Primary structural analysis of the transplantation antigens of the murine H-2 major histocompatibility complex. *Annu. Rev. Biochem.* 50:1025.
25. Brorson, K.A., S.W. Hunt III, T. Hunkapillar, 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.