

HUMAN LEUKOCYTE ANTIGEN F (HLA-F)
An Expressed HLA Gene Composed of a Class I Coding Sequence
Linked to a Novel Transcribed Repetitive Element

BY DANIEL E. GERAGHTY, XIAOHUA WEI,
HARRY T. ORR, AND BEVERLY H. KOLLER

*From the Department of Laboratory Medicine and Pathology, and the
Human Genetics Institute, University of Minnesota, Minneapolis, Minnesota 55455*

The class I genes of the human MHC influence many biologically important functions associated with the immune response. These genes encode highly polymorphic cell surface glycoproteins that represent the classical transplantation antigens HLA-A, -B, and -C found on almost all nucleated cells. These cell surface antigens participate in the formation of the molecular complex recognized by the antigen receptor of CTL (1, 2). As such, they are essential for the recognition and elimination of virus-infected cells by the immune system. A complete class I molecule consists of an α chain encoded by one of the closely linked polymorphic HLA-A, -B, -C genes noncovalently bound to a β chain encoded by the unlinked nonpolymorphic β_2 -microglobulin gene. Polymorphism of HLA class I antigens is believed to play an important role in establishing the diversity of immune responsiveness within populations and constitutes the major barrier to tissue transplantation.

The murine homologues of the human class I transplantation antigens are encoded by genes in the H-2 region and include the classical H-2 K, D, and L antigens. In addition to the genes encoding these antigens, the murine class I gene family includes an additional 18-26 genes that map to the Qa and Tla regions. The Qa- and Tla-encoded molecules are expressed on subpopulations of lymphoid cells and some leukemias (3-5). The differences in tissue distribution and the absence of extensive polymorphism suggest that the function of Qa and Tla molecules may differ from that of H-2 K, D, and L proteins.

In humans, there is considerable serological evidence for the existence of class I-like antigens whose expression suggests they represent the equivalents of the mouse Qa and Tla antigens (6, 7). Examination of human DNA with an HLA class I probe indicated that the HLA-A, -B, and -C genes represent only a small subset of an extensive gene family (8). Recent work cloning and characterizing all of the class I homologous sequences from a human lymphoblastoid cell line (LCL)¹ has resulted

This work was supported by grants from the National Institutes of Health, from the Leukemia Society, and by Institutional Cancer grant IN-26-30 from the American Cancer Society. B. H. Koller is a fellow of the Leukemia Society of America. B. H. Koller's present address is the Department of Pathology, University of North Carolina, Chapel Hill, NC 27599.

Address correspondence to D. E. Geraghty, Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, WA 98104.

¹ Abbreviations used in this paper: ICS, IFN consensus sequence; LCL, lymphoblastoid cell line; ut, untranslated.

in the identification of two class I-related genes (9-11). These genes have been sequenced and shown to encode intact HLA class I proteins distinct from the classical HLA-A, -B, and -C antigens. One gene, HLA-E, has been mapped between HLA-A and HLA-C and shown to be expressed in a variety of tissues. The second gene, designated HLA-6.0, encodes a class I-like molecule with a shortened cytoplasmic region. This latter gene is expressed on the cell surface after transfer of the cloned gene into human B-LCLs (12).

We describe here the cloning and characterization of a third human class I-related gene, HLA-5.4. The first six exons of HLA-5.4 have an organization similar to that of HLA-A2, and the extracellular and transmembrane domains of the protein encoded by this gene are similar to those of the HLA-A, -B, and -C antigens. The cytoplasmic segment of HLA-5.4 is predicted to be considerably shorter than that of other HLA class I molecules. The 3' untranslated region of HLA-5.4 is distinct from that of all other class I genes. The first 32 bp in this region are homologous to other class I genes while the remainder of this sequence diverges completely. Southern analysis indicated that the human genome contains 30-50 copies with homology to this unique 3' untranslated sequence, and Northern analysis showed that at least some of these related sequences are transcribed in a variety of tissues. In addition, we show that, unlike HLA-A, -B, and -C, the HLA-5.4 gene is differentially expressed. Message from the HLA-5.4 gene can be detected in resting T cells, skin cells, and in B-LCLs, but not in liver tissue or in the T cell line Molt-4.

Materials and Methods

Cloning of HLA-5.4. Genomic DNA was obtained from LCL 721.144 and digested to completion with the restriction enzyme Hind III. Digested DNA was size selected on a preparative agarose gel, and the fractionated DNA was eluted into an agarose/hydroxyapatite gel and purified as described (13). DNA from the fraction enriched for the 5.4-kb Hind III class I sequence was ligated to Hind III-cut dephosphorylated λ phage Charon 21a DNA, packaged in vitro, and transfected into *Escherichia coli* LE 392. Recombinants were screened by hybridization (14) with an HLA-B7 cDNA probe. 10 hybridizing plaques were picked, replated, and screened a second time. DNA prepared from the plaque-purified clones was digested with Hind III and analyzed by Southern blotting.

DNA Sequencing. A restriction map of pHLA-5.4 was constructed and appropriate fragments were subcloned into M13-mp19. For most of the sequence, sets of overlapping clones were isolated by the deletion method of Dale et al. (15). Where necessary, specific fragments were subcloned into M13-mp18 or M13-mp19. Sequencing reactions were carried out by the dideoxy-chain termination method (16). The entire sequence presented in Fig. 2 was determined on both strands, and regions that remained ambiguous were resolved by using deoxyinosine triphosphate as a substitute for deoxyguanosine triphosphate.

Analysis of RNA and DNA. RNA was prepared according to the guanidium isothiocyanate procedure described by Lizardi (17). T lymphocytes were isolated from heparinized peripheral blood of normal healthy donors by Ficoll-Hypaque density gradient centrifugation and nylon wool filtration. Cells were cultured in 1% PHA for 60 h before RNA isolation. Agarose gels containing formaldehyde were used to separate the mRNA, and blotting onto a nylon membrane and hybridization were carried out according to standard methods. Genomic DNA was isolated and digested with restriction enzymes according to the supplier's specifications. Digested DNA was separated on 0.7% agarose gels and transferred to a nylon membrane. Hybridization conditions have been described (18). Washing for RNA and DNA blots was in $0.1 \times$ SSC and 0.1% SDS at either 42°C or 68°C.

Preparation of Antisense RNA Probes and RNase Mapping. The 1,184-bp Bam HI-Pst I fragment from the 5' coding region of HLA-5.4 (positions 1,393-2,577 in Fig. 2) was prepared

and subcloned into the pSP64 vector according to standard methods. This subclone was characterized and used as a source of antisense RNA probe. Hybridization of the labeled probe (5×10^5 cpm) with 6 μ g total RNA was performed overnight in 80% formamide at 45°C (19). Protected fragments were analyzed on 6% acrylamide sequencing gels. pSP65 digested by Nci I and end labeled with polynucleotide kinase was used as size marker.

Computer Analysis of DNA and Protein Sequence. The DNA and deduced protein sequence of HLA-5.4 were analyzed with the Intelligenetics software on the SUN 3/160 computer of the Molecular Biology Computer Facility, University of Minnesota. Sequence homology searches were also performed on the Vax, University of Wisconsin.

Results

Isolation of the HLA-5.4 Gene. Southern blot analysis of LCL 721 genomic DNA had previously indicated that a number of non-HLA-A, -B, and -C class I genes were located within Hind III-generated fragments of 5.4 kb. Previous analysis of cosmid clones isolated from an LCL 721 cosmid library led to the cloning of one of these genes, HLA-5.4p. Subsequent examination of the HLA deletion mutant cell line 721.144 (20) with a probe specific for HLA-5.4p indicated that both alleles of HLA-5.4p were absent in this cell line. As 721.144 still contained a 5.4-kb Hind III fragment that hybridizes to a class I cDNA probe, this fragment must be contributed by a second non-HLA-A, -B, or -C locus.

To clone this remaining 5.4-kb class I sequence, a partial genomic library was constructed from size-selected Hind III-digested 721.144 DNA inserted in the λ phage vector charon 21a. Screening of the phage library with the HLA-B7 probe yielded 10 clones that all had identical restriction fragment patterns (data not shown). The 5.4-kb Hind III fragment contained in one of these clones was subcloned into puc19 and designated HLA-5.4. Fig. 1 shows a restriction map of the HLA-5.4 insert. With overlapping clones obtained by exonuclease digestion (15) and subcloning of restriction fragments, we determined 4,316 bp of the nucleotide sequence beginning at the 5' Pst I site and extending to the 3' Hind III site. Computer-assisted alignment revealed strong sequence homology with the HLA-A2 gene and predicted the assignment of exon-intron boundaries in HLA-5.4.

General Structure of the HLA-5.4 Gene. The HLA-5.4 DNA sequence and the deduced amino acid sequence are presented in Fig. 2. The overall structure of the HLA-5.4 gene is similar to other class I genes with sequence homology extending from the 5' flanking region into exon 8, the 3' untranslated sequence. There are, however,

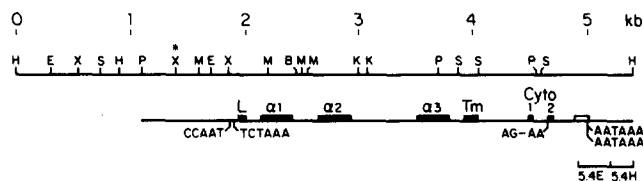


FIGURE 1. Restriction map of the 5.4-kb Hind III fragment containing the HLA-5.4 gene. B, Bam HI; E, Eco RI; H, Hind III; K, Kpn I; M, Sma I; P, Pst I; S, Sst I; X, Xba I. The asterisk marks the Xba I site present in the DNA sequence that is not cleaved by the enzyme

(see Discussion). A schematic diagram of the region sequenced is shown below. Exons are indicated by raised bars and labeled according to peptide domains. L, leader peptide; $\alpha 1$, $\alpha 2$, and $\alpha 3$, the three extracellular regions; Tm, transmembrane region; Cyto 1 and 2, the cytoplasmic regions. Control sequences and a variant 3' splice site dinucleotide are indicated. Portions of HLA-5.4 used as probes for Southern and Northern blots are indicated and designated 5.4E and 5.4H.

As is the case with HLA-A2, exon 8 of HLA-5.4 contains the last two nucleotides of the codon for the COOH-terminal amino acid, the stop codon, and the 3' untranslated region. After the stop codon, homology between HLA-5.4 and A2 continues for an additional 27 bp, after which the HLA-5.4 3' untranslated region diverges completely from that of any other class I gene. This unique sequence does contain two poly A addition signals (23) 90 bases downstream from the stop codon. Assuming that exon 7 is spliced out and that the message is polyadenylated using these signals, the HLA-5.4 mRNA is predicted to contain 1,250 nucleotides, including a 100-nucleotide poly A tail. This compares with a size of 1,600 nucleotides found for a similarly polyadenylated HLA-A, -B, or -C mRNA.

The unique portion of the HLA-5.4 3' untranslated region, from positions 3,799 to 4,316, was compared with other gene sequences compiled in the EMBO and Wisconsin gene banks. With the exception of several GT repeats found 3' of the poly A addition signals, no significant matches were found in either search.

Since this region was unique to HLA-5.4, a portion was subcloned for use as a locus-specific probe. The probe p54E extends from the Eco RV site at position 3,828 to the Taq I site at position 4,109. Fig. 3 shows Southern blot analysis of DNA from LCL 721 and from derived related cell lines. Under low stringency conditions (42°C), at least 30 distinct bands were seen. This suggested that the HLA-5.4 3' untranslated region contains a sequence homologous to a second distinct multigene family. This gene family did not appear to have been previously characterized since no homologies with published sequences could be found. Under high stringency conditions

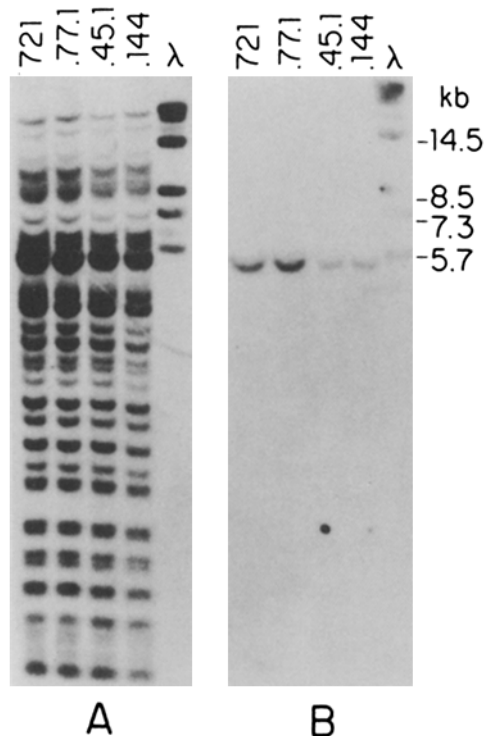


FIGURE 3. Southern analysis of DNA from LCL 721 and derived cell lines. 10 μ g of LCL 721 and mutant cell line DNA was digested with Hind III, separated on a 0.7% agarose gel, transferred to a nylon membrane, and hybridized with the subclone 5.4E. Mutant cell line .77.1 has a deletion of the short arm of chromosome 6 containing the A2:G1o1 haplotype; mutant .45.1 has a deletion of the short arm of chromosome 6 containing the A1:G1o2 haplotype; and .144 has a homozygous deletion of DNA at the HLA-A locus (20). The membrane was washed at 42°C in A and 68°C in B.

(68°C), p54E hybridizes to a single band at 5.4 kb, demonstrating that this probe is specific to the HLA-5.4 gene, and that the HLA-5.4 cloned sequence is representative of that found in the human genome.

Comparisons of the DNA and Protein Sequence of HLA-5.4. Comparisons of HLA-5.4 with five nonallelic class I genes are presented in Table I. As was found with HLA-E, the evolutionary position of HLA-5.4 is somewhat ambiguous. With the exception of intron 3 of HLA-E, HLA-5.4 has similar homology with any of these class I genes from the 5' flanking region through exon 5. The 3' portion of HLA-5.4 shows more homology to the HLA-A2 and -6.0 genes than to HLA-Bw58, -Cw3, or -E. Comparison of the sequences from intron 6 through intron 7 allows these six genes to be placed in two groups. The HLA-5.4, -A2, and -6.0 sequences share 86-90% homology among themselves, while sharing 71-79% homology with the HLA-Bw58, -Cw3, and -E sequences. Similarly, HLA-Bw58, -Cw3, and -E have 86-90% homology in this region.

The pattern of divergence of the HLA-5.4 protein sequence from other class I proteins is similar to that seen when comparing the HLA-A, -B, and -C genes (see Table I). The HLA-5.4 protein is highly conserved in the $\alpha 3$ domain presumably due to the importance of this region in the interaction with β_2 -microglobulin. The $\alpha 1$ and $\alpha 2$ domains show 62-76% homology with the corresponding domains of other HLA class I molecules. Similar to comparisons between HLA-A, -B, and -C proteins, the HLA-5.4 $\alpha 1$ and $\alpha 2$ domains are nearly equally diverged from those of HLA-A2, -Bw58, -Cw3, and -6.0 proteins. This is in contrast to HLA-E, which shows significantly more divergence in $\alpha 2$ than in $\alpha 1$.

Previous studies comparing class I protein sequences have identified hypervariable regions within the $\alpha 1$ and $\alpha 2$ domains (24). Most of the variable residues are clustered in regions from positions 57 to 83 in $\alpha 1$ and from positions 143 to 171

TABLE I
Percentage of DNA and Protein Sequence Identity between HLA-5.4 and Homologous Class I Genes

| Gene segment | 5.4/A2 | | 5.4/Bw58 | | 5.4/Cw3 | | 5.4/6.0 | | 5.4/E | |
|-----------------|--------|---------|----------|---------|---------|---------|---------|---------|-------|---------|
| | DNA | Protein | DNA | Protein | DNA | Protein | DNA | Protein | DNA | Protein |
| 5' Promoter | 80 | | 81 | | 79 | | 80 | | 75 | |
| Exon 1 | 85 | 86 | 88 | 72 | 90 | 86 | 92 | 81 | 78 | 71 |
| Intron 1 | 86 | | 88 | | 86 | | 89 | | 86 | |
| Exon 2 | 82 | 71 | 81 | 73 | 83 | 71 | 84 | 76 | 80 | 69 |
| Intron 2 | 81 | | 84 | | 83 | | 82 | | 82 | |
| Exon 3 | 81 | 71 | 84 | 76 | 84 | 76 | 82 | 73 | 78 | 62 |
| Intron 3 | 84 | | 85 | | 84 | | 86 | | 70 | |
| Exon 4 | 89 | 88 | 91 | 92 | 92 | 90 | 91 | 89 | 92 | 90 |
| Intron 4 | 83 | | 90 | | 85 | | 80 | | 90 | |
| Exon 5 | 93 | 82 | 87 | 74 | 88 | 72 | 91 | 82 | 91 | 77 |
| Intron 5 | 86 | | 82 | | 80 | | 88 | | 88 | |
| Exon 6 | 91 | 82 | 88 | 64 | 79 | 55 | 88 | - | 76 | 45 |
| Intron 6 | 87 | | 82 | | 72 | | 89 | | 81 | |
| Exon 7 | 87 | - | 83 | - | 75 | - | 79 | - | 79 | - |
| Intron 7 | 84 | | 68 | | 69 | | 86 | | 73 | |
| 3' untranslated | 92 | | 76 | | 79 | | - | | - | |

in $\alpha 2$. Likewise, the substitutions found in the HLA-5.4 protein are clustered within these regions (see Fig. 4).

Both of the cysteine pairs found in $\alpha 2$ and $\alpha 3$ and conserved among all MHC class I antigens are present in identical positions in the HLA-5.4 protein. A single *N*-linked glycosylation site, Asn-86, is found in the $\alpha 1$ region. The transmembrane segment of HLA-5.4 is identical in length and very similar in amino acid sequence to that of the HLA-A, -B, and -C proteins. The predicted cytoplasmic segment of HLA-5.4 would be 16 amino acids shorter than that of other HLA molecules if, as discussed above, exon 7 is spliced out of the mature HLA-5.4 transcript. The molecular mass of the HLA-5.4 protein should therefore be 2,000 daltons less than that of the classical HLA proteins. We have introduced the HLA-5.4 gene into B lymphoblastoid cells as part of an extrachromosomal plasmid (12). Expression of the HLA-5.4 protein in these cells could be demonstrated with mAb w6/32, a monomeric class I antibody, and mAb BBM.1, which binds free or α chain-associated $\beta 2$ -microglobulin. As predicted, the HLA-5.4 protein has an apparent molecular mass of 41,000 daltons, $\sim 2,000$ daltons less than that of the HLA-A, -B, and -C H chains.

While the general pattern of substitutions found in the HLA-5.4 protein is typical, an examination of the conserved residues of class I proteins shows HLA-5.4 to be significantly different. The recently determined three-dimensional structure of HLA-A2 and the delineation of the antigen binding site of class I antigens afford

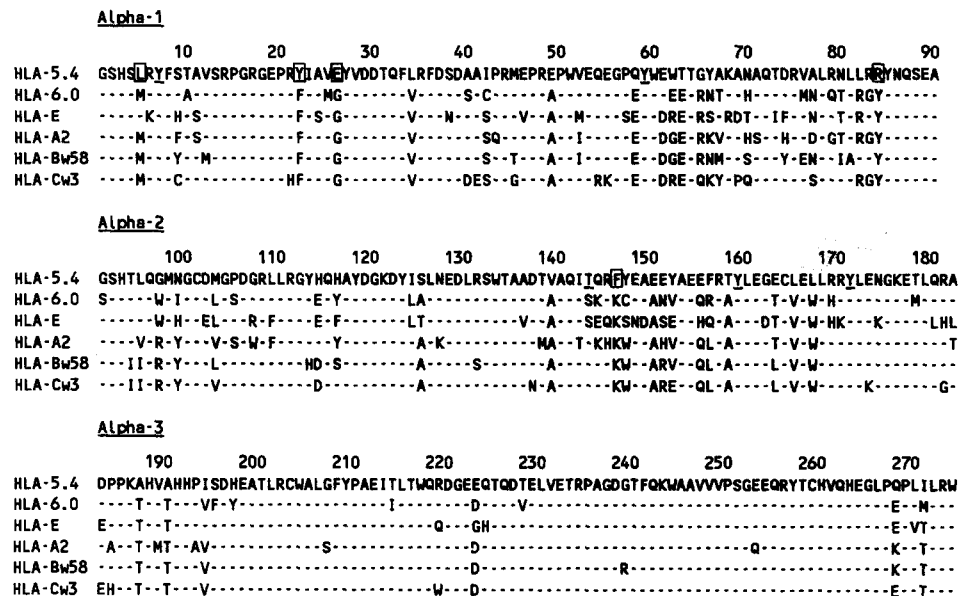


FIGURE 4. Comparison of the $\alpha 1$, $\alpha 2$, and $\alpha 3$ domains of class I HLA proteins. The amino acid sequence of the $\alpha 1$ - $\alpha 3$ domains of the HLA-5.4 protein is shown using the single letter amino acid code. Aligned underneath are corresponding sequences from HLA-6.0 (9), HLA-E (11), HLA-A2 (13), HLA-Bw58 (42), and HLA-Cw3 (43). Dashes denote identity to the respective position in the HLA 5.4 sequence. Amino acids boxed and underlined mark positions in the antigen binding site conserved among class I proteins (see Discussion).

an opportunity to examine the corresponding functional sites of the HLA-5.4 protein (25, 26). There are 10 completely conserved residues (including five tyrosines) pointing into the antigen recognition site in 22 human sequences. These residues in the HLA-5.4 protein are shown in Fig. 4. 5 of the 10 conserved residues are different in HLA-5.4. The significant changes include substitution of a tyr for phe at position 22, a glu for gly at position 26, and an arg for tyr at position 84. Four of the five conserved residues are tyrosines, and a fifth tyr residue is introduced at position 22. In contrast, 8 of the 10 residues are conserved in HLA-E, and 9 of the 10 are conserved in HLA-6.0, in both cases including all five tyrosines. In the murine Tla gene 37.1, 8 of the 10 residues are conserved, again including all five tyrosines (27). All of the substitutions at these positions in the latter three proteins are conservative.

The conserved residues in the long helix of the $\alpha 1$ domain are conserved in HLA-5.4, the only exception being an asp to glu substitution at position 61. The long helix of $\alpha 2$ shows more variability, including a lys to phe substitution at position 146 and an ala to glu substitution at position 150. The $\alpha 2$ helix in HLA-E shows approximately the same extent of divergence, with gly to asp substitution at position 162 and an arg to his at position 169.

Identification of Potential Promoter Sequences in the HLA-5.4 Gene. In the 5' untranslated region of class I genes, there are at least five promoter sequences thought to be important in regulating transcription. Sequences of class I promoter regions are shown aligned in Fig. 5, including the published sequences of HLA-A2 (13), HLA-E (11), and HLA-6.0 (9). The 5' 850 bp of HLA-5.4 are aligned with these other class I sequences, highlighting potential control sequences. The homology extends over the entire length of available sequence for any combination of the HLA-A2, HLA-6.0, and HLA-5.4 genes. The HLA-E gene shares homology for 190 bp 5' from the ATG start to homology A (see Fig. 5), at which point this sequence diverges completely from the others. The HLA-5.4 promoter has the CAAT and variant TCTAA homology found in most other HLA class I genes (an exception being HLA-E, which has the more typical TATAA homology).

Two potential regulatory sequences have been identified in murine class I genes (28) and have also been found in HLA class I genes (Fig. 5, A and B). Homology A is a 13-bp palindrome that has been shown to be the binding site for a protein factor that interacts with the H-2K promoter (29, 30), and a variant of this homology is also found in the murine β_2 -microglobulin promoter (30). Both the sequences from the H-2K and the β_2 -microglobulin promoters show enhancer and protein binding activities, while a sequence from the Q10 promoter, having two base substitutions relative to the H-2K homology, shows neither activity. Homology A in HLA-5.4 is altered at two positions, one of which, a G to A substitution at position three, is identical to that found in the Q10 sequence (28). As previously noted, the HLA-6.0 gene shows a deletion of part of this sequence, and the HLA-E gene has no sequence resembling this homology in its 5' region. Homology B is an 11-bp sequence found in all the published HLA-A, -B, and -C sequences, and a fragment containing a nearly identical sequence from the murine H2-K, -D, and L promoters has been shown to have enhancer activity (28). The HLA-5.4 promoter has a single base variant of this sequence. The substitution found in the HLA 5.4 homology B is distinct from either of the two substitutions found in the HLA-6.0 sequence and distinct from the single substitution found in the HLA-E homology.

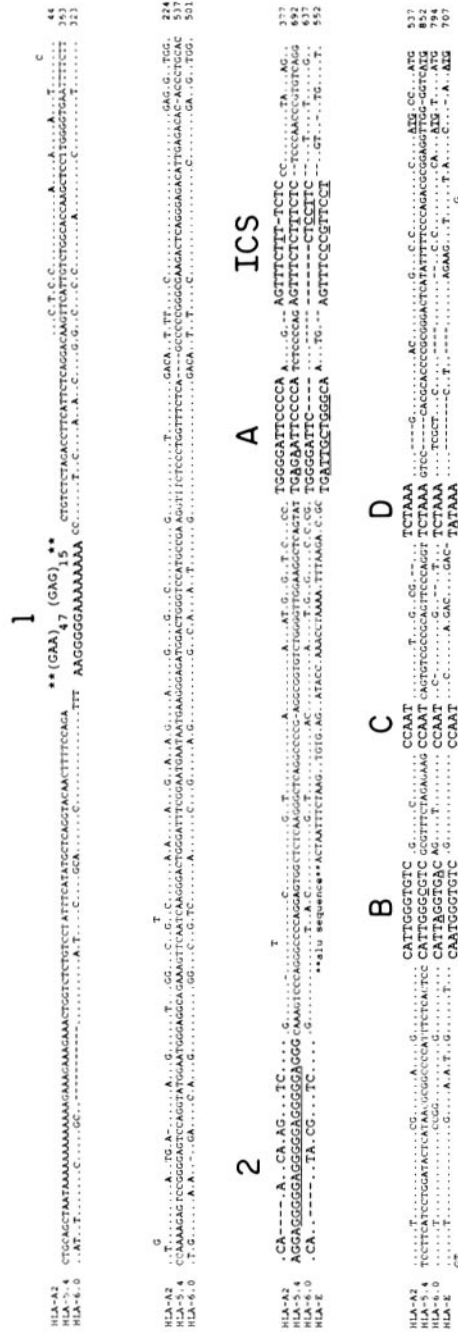


FIGURE 5. Comparison of the 5' promoter region of HLA-A2 (13), HLA-6.0 (9), and HLA-E (11) genes. Two polypurine tracts (1 and 2) and five enhancer and promoter sequences (A, B, C, D, and ICS) are shown in enlarged type. The first polypurine tract in HLA-5.4 is abbreviated. Nucleotides different from the HLA-5.4 sequence are shown. Dots indicate identity at that position and dashes indicate gaps introduced to achieve the best alignment. Bases shown above HLA-A2 or below HLA-E are insertions relative to HLA-5.4 and follow the base directly above or below. Bases different from the respective consensus sequence are underlined. The consensus for sequences A and B is from Kimura et al. (28); the core of the ICS is from Friedman and Stark (32), and the consensus for sequences C and D are from Koller and Orr (13). Also underlined are the ATG initiation codons. Numbers to the right refer to the last nucleotide on each line and correspond to the numbering used previously.

The IFN consensus sequence (ICS) represents the fifth putative promoter sequence found in the HLA class I genes (31, 32). An ICS very similar to the human ICS has also been found in a murine H-2L gene (33). As can be seen in Fig. 5, the HLA-6.0 gene has most of the ICS deleted and has no closely homologous sequence anywhere in the promoter region. The HLA-E 5' region contains a sequence with good homology to the ICS. The HLA-5.4 ICS is very close to the consensus and, in fact, is no more diverged than is the HLA-A2 ICS.

Examination of the HLA-5.4 promoter further upstream identifies two stretches of sequence similar to regulatory elements found near other mammalian genes (sequences 1 and 2, Fig. 5). Sequence 2 in HLA-5.4 consists of 25 purines with three tandem GGGGA repeats. Sequence alignment shows that HLA-A2 and HLA-6.0 are also purine rich in this region, while HLA-E is not. The HLA-A2 and -6.0 sequences are nearly identical to each other but are shorter, differ in base composition, and lack the repeat structure of the HLA-5.4 sequence. The most striking sequence to be found in the HLA-5.4 5' region is an uninterrupted stretch of 184 purine residues extending from positions 110 to 293 (Fig. 2). This region contains 47 occurrences of the trinucleotide GAA and 15 of GAG, including GAA tandemly repeated 21 times. Murine class I genes also contain a similar purine-rich stretch. The H-2K^k gene has a 45-bp region consisting of 42 purines with five GAAGA repeats ending 330 bp upstream of the ATG start (28). A closely homologous stretch of purines also occurs at the same position in the murine Qa region genes, Q10 and 27.1.

The HLA-6.0 gene does not contain a similar stretch of sequence. Despite close homology to HLA-5.4 on both the 5' and 3' sides of sequence 1, the corresponding HLA-6.0 sequence is much shorter (20 bp) and, while containing 15 consecutive purines, lacks any similar trinucleotide repeats. The HLA-Cw1 and -Cw2 genes both have a stretch of 10 purines, also without the trinucleotide repeat structure, in an analogous position (34). There is less homology between the surrounding sequences of these genes and those of HLA-5.4, however, making the alignment somewhat arbitrary.

Expression of the HLA-5.4 Gene and Related Transcripts. Northern analysis was carried out on a variety of human tissues and cell lines in order to determine which tissues express the HLA-5.4 gene and to examine whether any of the HLA-5.4 3' untranslated (ut) related sequences are expressed. Subclone p54E was used as the source of probe, since Southern blot analysis indicated that this sequence should hybridize with members of the HLA-5.4 3' related multigene family under low stringency conditions and specifically with HLA-5.4 under high stringency conditions. As can be seen in Fig. 6, the p54E probe hybridized with several distinct RNA species, some of which appear to be cell and tissue specific. One species, which appeared to be the most abundant message, was found in all tissues and cell lines examined. The relative size of this RNA was 600 nucleotides and, as seen in Fig. 6, it was polyadenylated.

The presence of the 600-nucleotide message in all tissues suggested that this mRNA might perform a housekeeping function and, therefore, that a similar RNA might be present in other mammalian tissue. To address this question, Northern and Southern analysis of mouse RNA and DNA was performed (Fig. 7). The HLA-5.4 ut multigene family probe hybridized to mouse DNA, and the complexity of the hybridization pattern suggested that these sequences comprise a similarly sized multigene family.

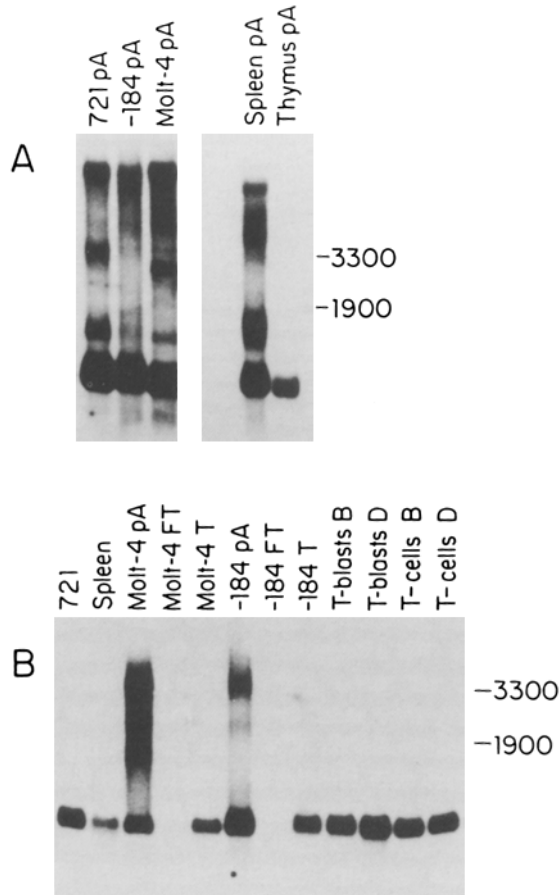


FIGURE 6. Northern analysis of RNA from various cell lines and tissues. RNA was fractionated on a 1% agarose gel containing 2.2 M formaldehyde, transferred to a nylon membrane, and hybridized with the 5.4E probe. pA, polyadenylated mRNA fraction; FT, flow-through nonpolyadenylated fraction; and T, total RNA. Washing was at 42°C. See Materials and Methods for details.

The Northern analysis showed that at least one of these mouse sequences is transcribed as an mRNA with size similar to the ubiquitous human message. Other mouse tissues and cell lines have been examined, and all showed a 600-nucleotide RNA species that hybridized to the p54E probe (data not shown).

In addition to the 600-nucleotide mRNA, there were at least four additional distinct RNA species that hybridized with p54E in LCL 721, Molt-4, and spleen cells. Not all of these RNA species were expressed in all cell types. For example, Molt-4 had a 2,800-nucleotide species not present in the other cell types, and RNA from spleen contained a 1,400-nucleotide species unique to this tissue. In addition, the relative amounts of RNA species common to these cells appeared to vary in different cell and tissue types. This could reflect differences in the level of expression of a given member of the multigene family. Alternatively, any single band may be comprised of more than one RNA species, some of which may be expressed in a tissue-specific manner.

To examine the expression of HLA-5.4, the Northern blots were subjected to the same high stringency washing used for the Southern analysis. However, it was not possible to unambiguously demonstrate expression of HLA-5.4 using this approach.

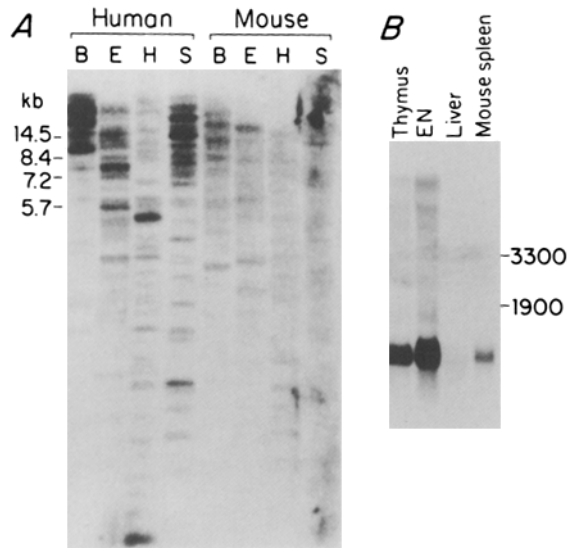


FIGURE 7. Northern and Southern analysis of human and mouse tissues. (A) Human DNA from LCL-721 and mouse DNA from cell line EL4 was digested to completion with the indicated enzymes (B, Bam HI; E, Eco RI; H, Hind III; S, Sst I). These DNAs were separated on an agarose gel, transferred to nylon membranes, and probed with the 5.4E probe. (B) Human RNA and mouse spleen RNA was fractionated on an agarose gel and subjected to Northern analysis with the 5.4E probe.

RNA was therefore analyzed using an RNase protection assay. The source of probe used in this analysis was a 1,184-bp Bam HI-Pst I fragment from the 5' coding region (positions 1,393-2,577) subcloned into the vector pSP64. This fragment contains all of exon 3, the HLA-5.4 coding sequence most diverged from other class I sequences. Of the full-length SP6-transcribed RNA, 274 bases protecting exon 3 of HLA-5.4 mRNA should be left undigested after RNase treatment. Fig. 8 shows protected fragments of the expected size could be found when hybridized to RNA from a variety of tissues and cell lines. T cells appear to have levels comparable with that found in B-LCLs and slightly higher levels than that found in PHA blasts. HLA-5.4 mRNA could also be detected in skin but not in RNA derived from liver. In addition, mRNA from B-LCLs derived from five unrelated individuals also protected a fragment of the same size (data not shown), implying that this portion of HLA-5.4 mRNA is not extensively polymorphic in the population. No protected fragment was seen when Molt-4 RNA was used, even after exposures 20-fold longer than that shown in Fig. 8. This analysis establishes that the HLA-5.4 gene is transcribed in human cells and indicates that this expression is tissue or cell type specific.

Discussion

HLA-5.4 is the third non-HLA-A,B,C gene isolated from LCL 721. The HLA-5.4 gene is unique in being composed of an HLA class I sequence linked to a sequence that is a member of a distinct multigene family comprising part of the 3' untranslated region. The sequence indicates that HLA-5.4 encodes an intact class I protein, and we have shown here that HLA-5.4 mRNA is present in B-LCLs and in resting T cells. Whereas HLA-E mRNA was found expressed at low levels in all cells and tissues examined, this was not the case with HLA-5.4, since no mRNA was found in the leukemic T cell line Molt-4.

Since the HLA-5.4 gene has a pattern of expression distinct from the five other class I genes, it might be expected that the HLA-5.4 promoter would contain unique

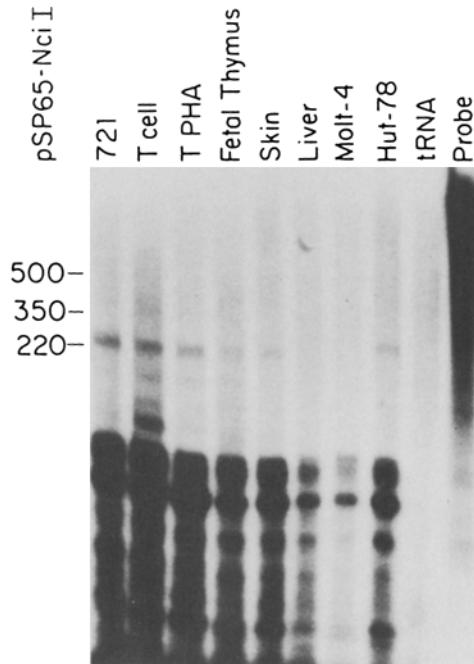


FIGURE 8. Expression of HLA-5.4 mRNA in different cell lines. The pSP64 probe and the RNase mapping assay used are described in Materials and Methods. The origin of the RNA is indicated above each lane. The undigested probe is included for comparison. The size marker is pSP65 cut with Nci I and end labeled.

regulatory elements. Studies of class I promoter sequences in the mouse have led to the identification and purification of a factor with binding activity specific for H-2 and β_2 -microglobulin enhancer sequences (30). The 13-bp palindromic sequence in the H-2K gene, to which this factor binds, is found unchanged in the HLA-A and -B genes. It is interesting to note that this sequence is found altered in the HLA-5.4 promoter and partially or completely absent in the HLA-6.0 and HLA-E promoters. The HLA-5.4 and HLA-6.0 genes are differentially expressed, and HLA-E is constitutively expressed at a much lower level than the HLA-A and -B genes. The HLA-C gene also contains an altered version of sequence A with two substitutions, both of which disrupt the palindromic nature of the sequence. HLA-C is expressed at $\sim 2\%$ the level of HLA-A and -B. It seems reasonable to suggest, then, that sequence A as found in the H-2K promoter and in HLA-A2 is necessary for the high level constitutive expression of HLA-A and -B genes.

Expression of the HLA-5.4 gene may be modified not only by the alteration of existing regulator motifs, but also by the presence of sequences not found in other class I genes. Thus, the short polypurine stretch found 300 bp upstream of the ATG codon (sequence 2 in Fig. 5) may have regulatory activity. Several studies have shown that S1 hypersensitive sites are frequently located within short polypurine/polypyrimidine tracts, and nuclear protein factors have been identified that bind at or near these sites (35). Some of these stretches, located in 5' flanking regions, can modulate local DNA structure when subjected to superhelical stress, thereby altering the conformation of neighboring sequences (36).

The second longer polypurine stretch (sequence 1 in Fig. 5) may also exert a unique regulatory influence on the HLA-5.4 gene. A similar size sequence with a different

repeat structure has been described in a chicken myosin H chain gene (37). The long polypurine stretch in the chicken myosin H chain gene appears to induce a continuous long range helical distortion of the flanking DNA. In this regard it is interesting to point out that the Xba I site found 5 bp 3' of the end of sequence 1 in HLA-5.4 is not cleaved by this enzyme. Since this site does not overlap with a dam methylation site, and is therefore presumably not methylated, we speculate that the helical distortion induced by sequence 1 may interfere with Xba I recognition or activity.

An additional aspect of sequence 1 concerns the means by which this sequence evolved. Since the HLA-5.4 and HLA-6.0 genes have arisen from a common ancestor, the manner in which these genes diverged points to different evolutionary mechanisms. The regions flanking sequence 1 appear to have diverged through random base substitutions. The polypurine stretch, on the other hand, must have evolved by DNA slippage and mispairing during replication and recombination. The HLA-6.0 sequence may have lost a part or all of a precursor sequence, while the HLA-5.4 sequence may have added sequence to the precursor. It has been suggested that ubiquitous slippage-like mechanisms are a major source of genetic variation, not predictable by the classical mutation process (38). Perhaps these genes represent examples of such mechanisms.

The predicted protein structure of HLA-5.4 suggests a biological role different from that of other class I proteins. Examination of the conserved residues has shown the HLA-5.4 protein is unique among all class I proteins, including examples from the murine Tla region. Only 5 of the 10 highly conserved residues pointing into the antigen binding site remain unaltered in HLA-5.4 (see Fig. 4). At least three of these substitutions are not conservative, implying that the function imparted on class I proteins by these residues has been lost or altered in HLA-5.4. Perhaps the HLA-5.4 protein is able to bind antigen but is restricted by a unique subset of TCRs. This possibility is intriguing considering the uncertainty of the ligand of the TCR- γ/δ .

A second question about the function of the HLA-5.4 protein relates to the pattern of cell surface expression. We have shown that HLA-5.4 mRNA is present in B-LCLs and that HLA-5.4 protein can be detected in the cytoplasm but not on the surface of these cells (12). One explanation is that HLA-5.4, like several murine Qa proteins, is secreted. This possibility seems unlikely because the transmembrane segment of the HLA-5.4 protein differs from the HLA-A2 protein at only four residues, and none of these changes would be expected to alter the hydrophobicity. The murine Q6 and Q7 proteins, both of which are secreted, also have a high degree of hydrophobicity in the TM regions, although both do contain single charged amino acid in the center of this segment (39). It is also possible that the shorter cytoplasmic segment predicted for HLA-5.4 might effect its membrane anchoring. However, studies of murine class I antigens with truncated cytoplasmic domains have shown that these altered proteins are expressed on the cell surface (40). In addition, HLA-6.0, which has a cytoplasmic segment of six amino acids, can be found on the cell surface (12). It is possible, however, that the HLA-5.4 protein is processed in some unknown manner resulting in its secretion. For the variable surface glycoprotein, which is known to be attached to cell membranes through a phospholipid tail, it has been demonstrated that the hydrophobic transmembrane region predicted from the cDNA sequence is absent from the mature protein (41).

One remarkable aspect of the HLA-5.4 gene is the sequence found in the 3' untranslated and flanking DNAs. Whether or not the presence of this sequence contributes to the unique expression pattern of HLA-5.4 is not known. The sequence is, however, a member of a multigene family consisting of at least 30 members. A number of these are transcribed either independently or as a result of their insertion into other expressed genes. The most abundantly transcribed sequence, a 600-nucleotide polyadenylated RNA, was found in all human cells and tissues examined. A similar number of related sequences were identified in the mouse genome, and at least one of these is transcribed, also yielding a 600-nucleotide mRNA. The function of this RNA is unknown but a comparison of the human and mouse sequences might be revealing. We have isolated cDNA clones derived from both mouse and human tissues (D. E. Geraghty and B. H. Koller, unpublished results) and are currently performing this analysis.

Several related transcripts distinct from the 600-nucleotide mRNA are also found in a variety of tissues, and some of these transcripts appear to be differentially expressed. It may be interesting to note that some of these transcripts have been found only in lymphoid tissues and derived cell lines. We have not yet determined whether the sequence shared by these transcripts is contained in the 3' untranslated region, as found in HLA-5.4, nor how much sequence is shared among them. Answers to these questions may reveal information about the unique function, if any, this sequence has as part of the HLA-5.4 mRNA.

The HLA-5.4 gene appears to represent a recombinant between members of two multigene families very diverse in sequence but with some general features in common. Both families are of similar size with the HLA class I family, containing 17 members compared with 30 sequences homologous to the p54E sequence. In addition, both have members that are transcribed constitutively at relatively high levels (the 600-nucleotide mRNA and HLA-A and -B), and members that are differentially expressed (the Molt-4 and spleen-specific mRNAs, HLA-6.0 and -E, and HLA-5.4, a member of both groups).

Summary

We describe here the isolation and sequencing of a previously uncharacterized HLA class I gene. This gene, HLA-5.4, is the third non-HLA-A,B,C gene characterized whose sequence shows it encodes an intact class I protein. RNase protection assays with a probe specific for this gene demonstrated its expression in B lymphoblastoid cell lines, in resting T cells, and skin cells, while no mRNA could be detected in the T cell line Molt 4. Consistent with a pattern of expression different from that of other class I genes, DNA sequence comparisons identified potential regulator motifs unique to HLA-5.4 and possibly essential for tissue-specific expression. Protein sequence analysis of human and murine class I antigens has identified 10 highly conserved residues believed to be involved in antigen binding. Five of these are altered in HLA-5.4, and of these, three are nonconservative. In addition, examination of the HLA-5.4 DNA sequence predicts that the cytoplasmic segment of this protein is shorter than that of the classical transplantation antigens. The 3' untranslated region of the HLA-5.4 gene contains one member of a previously undescribed multigene family consisting of at least 30 members. Northern analysis showed that several of these sequences were transcribed, and the most ubiquitous transcript, a

600- nucleotide polyadenylated mRNA, was found in all tissues and cells examined. This sequence is conserved in the mouse genome, where a similar number of copies were found, and one of these sequences was also transcribed, yielding a 600-nucleotide mRNA. The characterization of this unique HLA class I gene and the demonstration of its tissue-specific expression have prompted us to propose that HLA-5.4 be designated HLA-F.

Received for publication 27 April 1989 and in revised form 29 August 1989.

References

1. Goverman, J., T. Hunkapiller, and L. Hood. 1986. A speculative view of the multicomponent nature of T cell antigen recognition. *Cell*. 45:475.
2. Zinkernagel, R. M., and P. C. Doherty. 1979. MHC-restricted cytotoxic T cells: studies on the biological role of polymorphic major transplantation antigens determining T-cell restriction-specificity, function, and responsiveness. *Adv. Immunol.* 27:51.
3. Flaherty, L. 1981. The TLA region antigens. In *The Role of the Major Histocompatibility Complex in Immunobiology*. M. E. Dorf, editor. Garland Press, New York. 33-57.
4. Old, L. J., E. A. Boyse, and E. Stockert. 1963. Antigenic properties of experimental leukemias. I. Serological studies in vitro with spontaneous and radiation induced leukemias. *J. Natl. Cancer Inst.* 31:977.
5. Schlesinger, M., E. A. Boyse, and L. J. Old. 1965. Thymus cells of radiation chimeras: TL phenotype sensitivity to guinea pig serum, and origin from donor cells. *Nature (Lond.)*. 304:549.
6. Gazit, E., R. Gil, and R. Altshuler. 1987. Lymphocytes modulated by interferon express new non-HLA-A,B,C class I antigens. *Hum. Immunol.* 18:53.
7. van Leeuwen, A., M. J. Giphart, G. de Groot, B. Morolli, H. Festenstein, L. E. Nijenhuis, and J. J. van Rood. 1985. *Hum. Immunol.* 12:235.
8. Orr, H. T., F. H. Bach, H. L. Ploegh, J. L. Strominger, P. Kavathas, and R. DeMars. 1982. Use of HLA loss mutants to analyze the structure of the human major histocompatibility complex. *Nature (Lond.)*. 296:454.
9. Geraghty, D. E., B. H. Koller, and H. T. Orr. 1987. A human major histocompatibility complex class I gene that encodes a protein with a shortened cytoplasmic segment. *Proc. Natl. Acad. Sci. USA*. 84:9145.
10. Koller, B. H., D. Geraghty, H. T. Orr, Y. Shimizu, and R. DeMars. 1987. Organization of the human class I major histocompatibility complex genes. *Immunol. Res.* 6:1.
11. Koller, B. H., D. E. Geraghty, Y. Shimizu, R. DeMars, and H. T. Orr. 1988. HLA-E: a novel HLA class I gene expressed in resting lymphocytes. *J. Immunol.* 141:897.
12. Shimizu, Y., D. E. Geraghty, B. H. Koller, H. T. Orr, and R. DeMars. 1988. Transfer and expression of three cloned human non-HLA-A,B,C class I major histocompatibility complex genes in mutant lymphoblastoid cells. *Proc. Natl. Acad. Sci. USA*. 85:227.
13. Koller, B. H., and H. T. Orr. 1985. Cloning and complete sequence of an HLA-A2 gene: Analysis of two HLA-A alleles at the nucleotide level. *J. Immunol.* 134:2727.
14. Benton, W. D., and R. W. Davis. 1977. Screening lambda gt recombinant clones by hybridization to single plaques in situ. *Science (Wash. DC)*. 196:180.
15. Dale, R. M. K., B. A. McClure, and J. P. Houchins. 1985. A rapid single stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18S rDNA. *Plasmid*. 13:31.
16. Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe. 1980. Cloning is single-stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol.* 143:161.

17. Lizardi, P. M. 1983. Methods for the preparation of messenger RNA. *Methods Enzymol.* 96:24.
18. Koller, B. H., B. Sidwell, R. DeMars, and H. T. Orr. 1984. Isolation of HLA locus-specific DNA probes from the 3'-untranslated region. *Proc. Natl. Acad. Sci. USA.* 81:5175.
19. Little, P. F. R., and I. J. Jackson. 1987. Application of plasmids containing promoters specific for phage-encoded RNA polymerases. In *DNA Cloning*, Vol. III. D. M. Glover, editor. IRL Press, Washington, DC. 1-18.
20. Kavathas, P., R. DeMars, and F. H. Bach. 1980. Hemizygous HLA mutants of lymphoblastoid cells: a new cell type for histocompatibility testing. *Hum. Immunol.* 4:317.
21. Green, M. R. 1986. Pre-mRNA Splicing. *Annu. Rev. Genet.* 20:671.
22. Ruskin, B., J. M. Green, and M. R. Green. 1985. Cryptic branchpoint activation allows accurate *in vitro* splicing of human β -globin intron mutants. *Cell.* 41:833.
23. Manley, J. L. 1988. Polyadenylation of mRNA precursors. *Biochim. Biophys. Acta.* 950:1.
24. Lopez de Castro, J. A., J. A. Barbosa, M. S. Krangel, P. A. Biro, and J. L. Strominger. 1985. Structural analysis of the functional sites of class I HLA antigens. *Immunol. Rev.* 85:149.
25. Bjorkman, P. J., M. A. Saper, B. Samraoui, W. S. Bennett, J. L. Strominger, and D. C. Wiley. 1987. Structure of the human class I histocompatibility antigen, HLA-A2. *Nature (Lond.).* 329:506.
26. 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.
27. Transy, C., S. R. Nash, B. David-Watine, M. Cochet, S. W. Hunt III, L. E. Hood, and P. Kourilsky. 1987. A low polymorphic mouse H-2 class I gene from the Tla complex is expressed in a broad variety of cell types. *J. Exp. Med.* 166:341.
28. Kimura, A., A. Israel, O. Le Bail, and P. Kourilsky. 1986. Detailed analysis of the mouse H-2K^b promoter: enhancer-like sequences and their role in the regulation of class I gene expression. *Cell.* 44:261.
29. Baldwin, A. S., Jr., and P. A. Sharp. 1987. Binding of a nuclear factor to a regulator sequence in the promoter of the mouse H-2K^b class I major histocompatibility gene. *Mol. Cell. Biol.* 7:305.
30. Israel, A., A. Kimura, M. Kieran, O. Yano, J. Kanellopoulos, O. Le Bail, and P. Kourilsky. 1987. A common positive trans-acting factor binds to enhancer sequences in the promoters of mouse H-2 and β_2 -microglobulin genes. *Proc. Natl. Acad. Sci. USA.* 84:2653.
31. Burrone, O. R., and C. Milstein. 1982. Control of HLA-A,B,C synthesis and expression in interferon-treated cells. *EMBO (Eur. Mol. Biol. Organ.) J.* 1:345.
32. Friedman, R. L., and G. R. Stark. 1985. α -Interferon-induced transcription of HLA and metallothionein genes containing homologous upstream sequences. *Nature (Lond.).* 314:637.
33. Sugita, K., J. Miyazaki, E. Appella, and K. Ozato. 1987. Interferons increased transcription of a major histocompatibility class I gene via a 5' interferon consensus sequence. *Mol. Cell. Biol.* 7:2625.
34. Gussow, D., R. S. Rein, I. Meijer, W. de Hoog, G. H. A. Seemann, F. M. Hochstenbach, and H. L. Ploegh. 1987. Isolation, expression, and the primary structure of the HLA-Cw1 and HLA-Cw2 genes: evolutionary aspects. *Immunogenetics.* 25:313.
35. Collier, D. A., J. A. Griffin, and R. D. Wells. 1988. Non-B right-handed DNA conformations of homopurine. Homopyrimidine sequences in the murine immunoglobulin C switch region. *J. Biol. Chem.* 263:7397.
36. Kohwi-Shigematsu, T., and Y. Kohwi. 1985. Poly(dG)-Poly(dC) sequences, under torsional stress, induce an altered DNA conformation upon neighboring DNA sequences. *Cell.* 43:199.

37. McCarthy, J. G., and S. M. Heywood. 1987. A long polypyrimidine/polypurine tract induces an altered DNA conformation on the 3' coding region of the adjacent myosin heavy chain gene. *Nucleic Acids Res.* 15:8069.
38. Tautz, D., M. Trick, and G. A. Dover. 1986. Cryptic simplicity in DNA is a major source of genetic variation. *Nature (Lond.)* 322:652.
39. Stroynowski, I., J. Forman, R. S. Goodenow, S. G. Schiffer, M. McMillan, S. O. Sharrow, D. H. Sachs, and L. Hood. 1985. Expression and T cell recognition of hybrid antigens with amino-terminal domains encoded by Qa-2 region of major histocompatibility complex and carboxyl termini of transplantation antigens. *J. Exp. Med.* 161:935.
40. Goodenow, R. S., I. Stroynowski, M. McMillan, M. Nicolson, K. Eakle, B. T. Sher, N. Davidson, and I. Hood. 1983. Expression of complete transplantation antigens by mammalian cells transformed with truncated class I genes. *Nature (Lond.)* 301:388.
41. Cross, G. A. M. 1987. Eukaryotic protein modification and membrane attachment via phosphatidylinositol. *Cell.* 48:179.
42. Ways, J. P., H. L. Coppin, and P. Parham. 1985. The complete primary structure of HLA-Bw58. *J. Biol. Chem.* 260:11924.
43. 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.