MOLECULAR AND CELLULAR BIOLOGY, Jan. 1990, p. 243–253
0270-7306/90/010243-11$02.00/0

Vol. 10, No. 1

Transcription Analysis, Physical Mapping, and Molecular Characterization of a Nonclassical Human Leukocyte Antigen Class I Gene

MICHAEL J. CHORNEY,† IKUHISA SAWADA, GERALD A. GILLESPIE, RAKESH SRIVASTAVA, JULIAN PAN, AND SHERMAN M. WEISSMAN*

Department of Human Genetics, Yale University School of Medicine, New Haven, Connecticut 06510

Received 24 May 1989/Accepted 25 September 1989

The human major histocompatibility complex contains approximately 20 class I genes, pseudogenes, and gene fragments. These include the genes for the three major transplantation antigens, *HLA-A*, *HLA-B*, and *HLA-C*, as well as a number of other genes or pseudogenes of unknown biological significance. Most of the latter have C+G-rich sequences in their 5′ ends that are unmethylated in the B-lymphoblastoid cell line 3.1.0. We investigated one of these genes, *HLA-H*, in more detail. The gene is, overall, strongly homologous in sequence to *HLA-A* but differs in several potentially significant ways, including changes in conserved promoter sequences, a single-base deletion producing a translation termination codon in exon 4, and a region of sequence divergence downstream of the transcribed portion of the gene. Nevertheless, mouse L cells transfected with the gene accumulated small amounts of apparently full-length polyadenylated RNA. A portion of this RNA begins at the transcription site predicted by analogy to certain class I cDNA clones, while another portion appears to begin shortly upstream. L cells transfected with a hybrid gene containing the first three exons of *HLA-H* and the last five exons of *HLA-B27* accumulated full-length *HLA* transcripts at the same level as cells transfected with an *HLA-B27* gene; both levels are at least 15- to 20-fold higher than that directed by *HLA-H* alone. In addition, we isolated a cDNA clone for *HLA-H* that contains a portion of intron 3 attached to a normally spliced sequence comprising exons 4 through 8. These results suggest that low levels of translatable mRNA for the truncated class I heavy chain encoded by *HLA-H* are produced under physiologic circumstances and that sequences 3′ of intron 3 decrease the levels of stable transcripts.

The human major histocompatibility complex (MHC), located on the short arm of chromosome 6, possesses a diverse array of genes, some of which are intimately involved in the immunological presentation of foreign antigen to immunocytes (29). Those genes that encode proteins of approximately 40 kilodaltons which are in noncovalent association with β2-microglobulin are termed human leukocyte antigen (HLA) class I genes. In addition to encoding proteins capable of binding β2-microglobulin, all class I genes share a common organization consisting of eight exons. Exon 1 encodes the leader sequence; exons 2, 3, and 4 encode the three external domains of each transplantation antigen; exon 5 encodes a membrane anchor domain; and exons 6 and 7 and a portion of exon 8 encode the regions of the protein which interact with the membrane and cytoplasm (25).

In humans, there are approximately 20 class I genes, pseudogenes, or gene fragments (27, 32, 69) contained within nearly 2,000 kilobases (kb) of DNA (9, 44, 60, 61). The best-studied ones are those termed *HLA-A*, *HLA-B*, and *HLA-C*, which encode proteins expressed on virtually all somatic cells. These proteins bind foreign peptides which, upon reaching the cell surface, interact with syngeneic cytotoxic T lymphocytes that initiate restrictive killing of these infected targets. Aside from these three genes, which are among the most polymorphic in the human genome, little is known about the diverse, expressible members of this family, such as *HLA-RS5* (68); reexamination of the sequence of the *HLA-RS5* gene has revealed that it differs by less than 2% from *HLA-E* and is therefore allelic to this gene [J. Pan, J. Germino, M. Chorney, and S. Weissman, Proc. Natl. Acad. Sci. USA **86**:650, 1989]), *HLA-E* (34, 55), and *HLA-6.0* (22).

BALB/c and B10 mice have more class I genes than do humans. These are found predominantly in the *Qa* and *TL* regions (75, 77), both of which are telomeric to the classical transplantation antigen genes contained within the *H-2* region. The genes of the *Qa* region are characterized by in-frame termination codons found within their exon 5 sequences. These lead to production of shortened proteins lacking cytoplasmic tails (20, 74). Most of the proteins encoded by *Qa* genes are bound to the surfaces of hepatocytes and other cells through an aspartate phosphatidylinositol linkage (73). One *Qa* gene, *Q10*, encodes a β2-microglobulin-associated secretory class I protein whose function is not known (38).

The possible presence of genes analogous to *Qa* genes within the human MHC is currently under examination. The recent sequencing of gene *HLA-6.0* (22), which contains an uncharacteristic termination in exon 6, supports the possibility that analogous genes do exist. Like their possible mouse counterparts, most nonclassical class I genes are reported to map to the telomeric portion of the MHC (58). One such gene, which we have provisionally termed *HLA-H*, maps within a maximum of 150 kb of *HLA-6.0*, as demonstrated by chromosomal hopping (M. J. Chorney, H. Shukla, G. A. Gillespie, I. Sawada, G. Carle, F. Collins, and S. M. Weissman, manuscript in preparation). This HLA class I gene or pseudogene contains an in-frame termination within exon 4 resulting from a single-nucleotide deletion. In

* Corresponding author.
† Present address: Department of Microbiology and Immunology, The Pennsylvania State University College of Medicine, The Milton S. Hershey Medical Center, Hershey, PA 17033.

this report, we present the complete nucleotide sequence and molecular mapping of *HLA-H*; we also describe its expression at the transcriptional level in transfected mouse L cells and a human B-lymphoblastoid line.

MATERIALS AND METHODS

Cells. The MHC hemizygous B-lymphoblastoid cell line 3.1.0 (*HLA-A2 HLA-B27 HLA-CW1*) (45), was maintained in RPMI 1640 supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin. Mouse L cells previously transfected with human β2-microglobulin and herpes simplex virus thymidine kinase genes were maintained in Dulbecco modified Eagle medium supplemented with 10% horse serum–1% penicillin-streptomycin–hypoxanthine-aminopterin-thymine. This cell line, known as J27, was kindly supplied by Paula Kavathas (Yale University, New Haven, Conn.). Mouse-6 (original designation, MCH 262A1/P6), which is a hybrid cell line composed of human chromosome 6 on a mouse background, was maintained in Dulbecco modified Eagle medium–10% fetal bovine serum–1% penicillin-streptomycin–G418 (600 g/ml). These cells were a kind gift of Eric Stanbridge (University of California, Irvine). The human lines Molt-4, Hut102, and KMOE were grown in RPMI 1640–10% horse serum–1% penicillin-streptomycin.

Transfections. Transfections were performed as described by Wigler et al. (76), except that 25 μg of cosmid DNA and 2.5 μg of pSV2neo DNA were used per experiment. High-molecular-weight carrier DNA was omitted. After selection in G418 (300 μg/ml)-containing Dulbecco modified Eagle medium–hypoxanthine-aminopterin-thymine, Neo^r J27 colonies (which numbered greater than 1,000 per experiment) were pooled for future analysis.

Libraries and gene isolations. The *HLA-H* class I gene was originally isolated from the Charon 4a placenta library of Maniatis et al. (48) as previously described (7). Its previous designation was *LN11A*. The 6.6-kb *Eco*RI fragment carrying this gene was subsequently subcloned into pBR328 to facilitate sequencing analysis. Multiple overlapping clones containing *HLA-H* were also isolated from 3.1.0 cosmid libraries by using *HLA-B7* cDNA as a hybridization probe. Two cosmid libraries were screened. One contained partially *Sau*3A-digested, size-fractionated DNA ligated into the *Bam*HI site of cosmid vector pcos2EMBL; the other contained partially *Eco*RI-digested, size-fractionated DNA ligated into the *Eco*RI site of cosmid vector pJB8. Production of the cosmid libraries is described in reference 43. The 6.6-kb *Eco*RI fragment carrying *HLA-H* from 3.1.0 cosmid 6 (CosRS6) was subcloned into the Bluescript KS(−) plasmid vector to facilitate further analysis and termed pHLA-H. *HLA-A2*-containing cosmid 28 (CosRS28) from the 3.1.0 library was identified by using the 3′ *HLA-A2*-specific probe of Koller et al. (36). The *HLA-B27* cosmid (CosRS39) was identified by using the 544-base-pair *Pvu*II fragment derived from the *HLA-B7* gene 3′-untranslated region (12) as a probe. Further confirmation that the gene was *HLA-B27* was supplied by transfection. J27 cells transfected with CosRS39 reacted with HLA-B7, HLA-B27, and HLA-Bw22-specific monoclonal antibody ME1 (17). The 6.6-kb *Eco*RI fragment containing the HLA-*B27* gene from CosRS39 was subcloned into pUC19 and termed pB27.

The 3.1.0 cDNA library was produced from poly(A)$^+$ RNA with a lambda ZAP vector system (Stratagene) by standard methods (4, 49). Single-stranded cDNA contained in the Bluescript plasmid vector was induced by using M13 helper bacteriophage R408 according to Stratagene instructions.

DNA sequencing. DNA fragments to be sequenced were subcloned into M13 vectors mp18 and mp19 (57) for single-stranded sequencing or plasmid vector Bluescript KS(−) for double-stranded sequencing. Single-stranded DNA templates or denatured double-stranded DNA plasmid subclones were sequenced by the dideoxynucleotide chain termination method (64) with [^{35}S]dATP and Sequenase (United States Biochemical). In areas where compression was observed, the sequencing reactions were performed with either dITP or 7-deaza-dGTP nucleotides.

Cosmid mapping. 3.1.0 overlapping cosmid clones CosRS6 and CosRS42, which contain the *HLA-H* gene, and CosRS28, which contains the *HLA-A2* gene, were linearized with either *Cla*I or *Sal*I and subsequently partially digested with *Eco*RI. The digests were then electrophoresed at room temperature on a field inversion vertical gel (1% agarose, 0.5% TBE [0.089 M Tris borate, 0.089 M boric acid, 0.002 M EDTA; pH 8.0]; 150 V) and blotted onto a Biotrans nylon filter. Electrophoresis forward time was 0.5 s, and reverse time was 0.25 s, as described in reference 8. The filters containing partial digests of CosRS42 and CosRS28 were probed consecutively with the left (α)- and right (β)-end *Bam*HI-*Eco*RI fragments derived from digestion of the pcos2EMBL vector (0.4- and 1.8-kb fragments, respectively). Filters containing complete *Sal*I and partial *Eco*RI digests of CosRS6 DNA were probed with both *Eco*RI-*Sal*I pJB8 vector fragments. To reuse the blots after autoradiography, the bound probe was removed by boiling the filters in distilled water for 5 min (Northern (RNA) blots were treated the same way to remove the radioactive probe). High-molecular-weight and 1-kb ladder DNA standards (Bethesda Research Laboratories, Inc.) were run on all gels for determination of molecular size.

Pulsed-field gel electrophoresis. 3.1.0 DNA was cleaved in agarose blocks as previously described (65). The restricted fragments in agarose plugs were run on orthogonal field alteration gel electrophoresis (OFAGE) gels with an LKB Pulsaphor gel apparatus with a pulse time of 2 min as described in reference 65. The electrophoresis run time was 40 h. After electrophoresis, the gels were treated with 0.2 N HCl for 15 min, followed by denaturation and neutralization. The gels were then blotted against 20× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) onto a Hybond nylon filter (Amersham Corp.), vacuum dried, and UV irradiated for 2 min before prehybridization.

RNA isolation and RNase protection. Total cytoplasmic RNA was isolated by the Nonidet P-40 lysis method of Favaloro et al. (18). For Northern blotting, RNA was electrophoresed on 1.1% agarose gels containing formaldehyde and blotted directly with 20× SSPE (3 M NaCl, 0.25 M $NaH_2PO_4 \cdot H_2O$, 0.04 M EDTA [pH 7.4]) onto Biotrans nylon filters.

RNase protection assays were performed as described by Melton et al. (54). As a probe, the 1.2-kb *Eco*RI-*Pst*I fragment spanning exon 1 and containing the entire 5′ end of the *HLA-H* gene was subcloned into pBluescript KS(−). Before labeled RNA probes were generated from this template with T7 RNA polymerase, the subcloned DNA was linearized with *Hin*dIII. The labeled RNA probe of 838 nucleotides was purified on a 6% acrylamide–urea gel. For each RNA sample to be analyzed, 10^5 cpm of the probe was added to approximately 20 μg of total cytoplasmic RNA. Hybridization was allowed to proceed overnight at 53°C. After hybridization, the solutions were treated with RNase A and RNase T1, followed by electrophoresis on a 6% acrylamide–urea gel.

Hybrid gene construction. pHLA-H and pB27 were both cleaved with *Kpn*I. This enzyme digests both genes at their exon 3-intron 3 splice junctions. In addition, *Kpn*I cleaves both plasmid sequences in the polycloning site located 3′ of each gene. Isolation of the 3′ *Kpn*I fragments and religation to the opposite *Kpn*I fragments containing the 5′ ends of the genes, as well as plasmid sequences, completed the reconstruction of hybrid genes (see Fig. 7).

Nucleic acid blotting and hybridization. DNA blotting was performed by the method of Southern (67), while filter hybridization was performed by the method of Hanahan and Meselson (23). The prehybridization and hybridization solutions used, including those used for Northern blotting, contained 0.3% nonfat dried milk, as described by Johnson et al. (26). Blots of pulsed-field gels were hybridized by the method of Church and Gilbert (11).

All probes used for hybridization were labeled by the random-hexamer-priming method of Feinberg and Vogelstein (19) to a specific activity of $>10^8$ cpm/μg.

Computer analysis. All sequencing data were manipulated with the software package supplied by the University of Wisconsin Genetics Computer Group (14) on the Yale University School of Medicine Vax computer.

RESULTS

During our investigations into the possible expression of 3.1.0 MHC class I genes after transfection into mouse L cells, we observed transcripts associated with the *HLA-A2*, *HLA-B27*, and *HLA-Cw1* genes which encode the classical transplantation antigens, as well as with certain other uncharacterized class I genes. One such line, transfected with cosmid clone RS5, stimulated further study which ultimately led to identification of a nonclassical but expressible HLA class I gene (68). This gene has been termed *HLA-E* (2) and has been reported to be expressed most abundantly in resting T lymphocytes (34). Like RS5, the class I gene contained in the overlapping 3.1.0 cosmid clones CosRS6 and CosRS42 also gave rise to RNA transcripts after transfection into J27 cells at a level slightly but reproducibly above the background (M. Chorney, H. Vasavada, and S. Ganguly, unpublished data).

We therefore proceeded to further study the gene contained within the 6.6-kb *Eco*RI fragment of CosRS6, as well as to reinvestigate the nucleotide sequence of *LN11A*, a previously described class I gene derived from placental DNA (7, 66). The latter was undertaken because the *LN11A* gene and the class I gene contained in CosRS6 yielded identical double-digest patterns after cleavage with all of the restriction endonucleases tested. Reexamination and comparison of the sequence of pLN11A with that of the gene contained in CosRS6 showed that the sequences were virtually identical, differing by less than 0.1%. The base differences between these genes were found only in the noncoding sequences. We present in Fig. 1 the complete nucleotide sequence of the 6.6-kb *Eco*RI fragment; this sequence is compared to the sequence of *HLA-A2* derived from 3.1.0. Because of the characteristics of these genes and the pulsed-field mapping results presented below, we have retermed *LN11A* and its alleles *HLA-H*. This is in keeping with the recent publication of HLA designations F and G for class I genes *HLA-5.4* and *HLA-6.0*, respectively (33).

The structure of *HLA-H* is similar to that of other class I genes, consisting of eight exons homologous in sequence to those of the classical transplantation antigen genes. The sequence of exons 1, 2, and 3 of *HLA-H* could encode protein domains whose structure is very similar to that of known functional class I genes. However, exon 4 of both the 3.1.0 gene and the gene from the placental library showed on repeated sequencing a frameshift resulting from a decrease of 1 base in the length of this exon. Downstream of the deletion is an in-frame termination codon (UGA). The consequence of this frameshift and subsequent termination is that an mRNA of *HLA-H*, if spliced in the same pattern as that which occurs with other MHC class I genes, would encode a protein that terminates just before the end of exon 4. This protein would lack transmembrane and cytoplasmic segments and would have a molecular weight of approximately 30,000.

The promoter region of *HLA-H* showed several differences in regions where the sequence is conserved in other class I genes. For example, the consensus *H-2* sequence TGGGGATTCCCCA within enhancer A (3, 28) is replaced by TGGGGATTCGCCA at residues 997 through 1009 in *HLA-H*, while the *H-2* consensus sequence CATTGGGTG GC within the enhancer B sequence is replaced by CACTAG GTGTC (three nucleotide changes) at residues 1092 through 1102. *HLA-A2* is identical to both of the consensus sequences, except for the same G-to-T change in the enhancer B sequence as in *HLA-H*.

Overall, *HLA-H* is more homologous to *HLA-A* genes, particularly *HLA-A2*, than to *HLA-B* genes. This similarity is evident when both the 5′-flanking sequences and exons 8 are compared and extends to the 3′ end of the 6.6-kb fragment (90.8% homology); however, excluded from the comparison are residues 5742 to 5824 3′ of the *HLA-H* gene and residues 5755 to 5966 3′ of *HLA-A2*, where there is no significant homology in the sequences (each insertion or deletion, even when composed of multiple nucleotides, was considered as a single change) (Fig. 1). When nucleotide changes were compared only within the coding regions (up to the TGA termination in exon 8 of *HLA-A2*; 1,092 positions were considered), 91.0% homology was observed. Of the 98 nucleotide differences found within the coding regions, 37.8% were at position 3. The sequence of exon 8, which contains untranslated nucleotides, is characteristic for each group of allelic class I genes (39); in a comparison with *HLA-A2* and *HLA-B27*, the nucleotide homologies of exon 8 of *HLA-H* were approximately 93.2 and 84.2%, respectively. The overall nucleotide homology of the *HLA-H* gene and its flanking sequences compared with that of pseudogene *HLA-12.4* (47) was greater than to any other class I gene, approximately 98% in the region where comparison was possible. However, the *HLA-12.4* gene does not possess a deleted nucleotide in exon 4. This homology suggests either that *HLA-12.4* is an allele of *HLA-H* or that the genes arose as the result of a very recent gene duplication.

The approximately 91% nucleotide sequence homology of *HLA-H* to *HLA-A2* extends through both the 5′- and 3′-flanking regions (approximately 6.5 kb of the sequence was included in the comparison). To determine whether this homology extends into the surrounding DNA, we isolated the *HLA-A2*-containing cosmid from a 3.1.0 cosmid library by using the 6.6-kb *Eco*RI fragment carrying the *HLA-H* gene as a probe. We confirmed that the gene in CosRS28 was *HLA-A2* by using the *HLA-A2*-specific probe of Koller et al. (36). This probe strongly hybridized to a 6.7-kb *Eco*RI fragment, even after extensive high-stringency washing. The subclone of this fragment, after transfection into J27 cells, directed the synthesis of a 40-kilodalton cell surface protein reactive with *HLA-A2*-specific monoclonal antibody MA2.1 (53) (data not shown). The sequence of the gene presented in

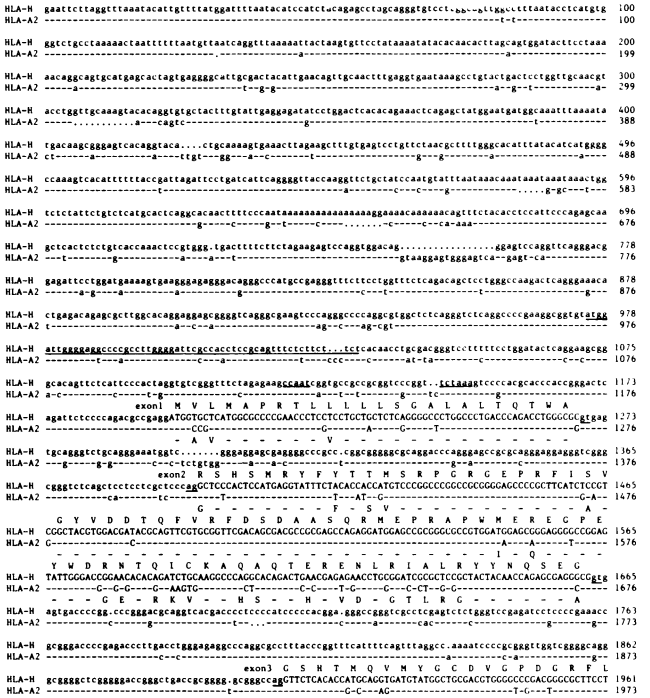

FIG. 1. Complete nucleotide sequence of the 6.6-kb *Eco*RI fragment carrying the *HLA* class I gene *HLA-H*; comparison to the nucleotide sequence of the 6.7-kb *Eco*RI fragment containing the *HLA-A2* gene from the 3.1.0 cell line. Exons are displayed in uppercase letters; introns and flanking sequences are displayed in lowercase letters. Nucleotide matches in the *HLA-A2* sequence are indicated by minus signs. Sequence alignments were determined by using the GAP program provided by the University of Wisconsin Genetics Computer Group. The translation of nucleotides within exons 1 through 8 for both *HLA-H* and *HLA-A2* is in the uppercase single-letter codes. The single-nucleotide deletions in exons 4 and 7 of *HLA-H* (see text) were ignored to maximize the amino acid homology to *HLA-A2*. Both of these deletions and the two nucleotides preceding them are marked by boxes. If one considers the exon 4 deletion, the shifted reading frame of exon 4 becomes (starting with CAC) HTRSSWRPGLQGLQGMEPSRSGRLWWCLLER SRDTPAMCSMRVCQSPSPTer. Underlined, starting from the top, are the sequences for enhancer A and the interferon response elemen., RNA polymerase II promoter sites (ccaat and tctaaa), RNA splice acceptor sites (ag/gt), and the polyadenylation signal at position 4479.

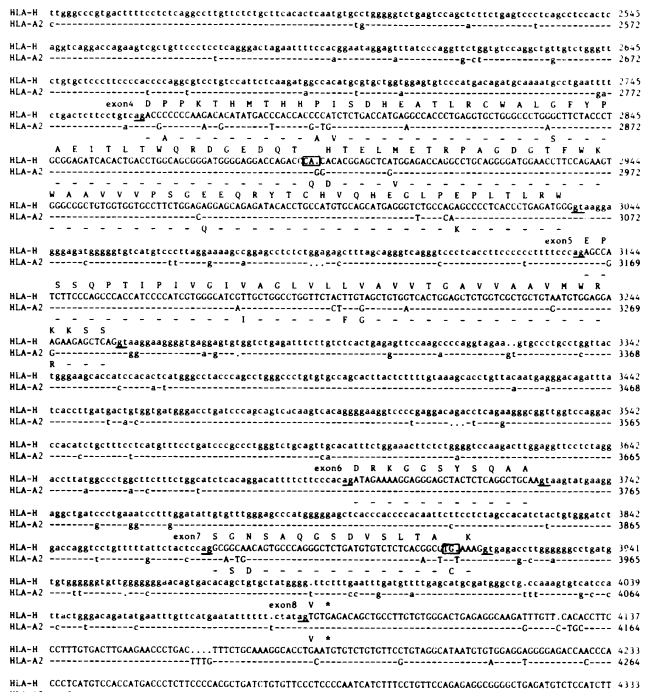

Fig. 1 was found to diverge no more than 0.1 to 0.2% from the *HLA-A2* gene sequenced by Koller and Orr (35) in the regions sequenced by them.

Figure 2 shows the *Eco*RI restriction maps of CosRS28 and CosRS6. The structure of CosRS28 is broadly similar to that of the *HLA-A3* and *HLA-Aw24* cosmids reported previously by N'Guyen et al. (56). All of the *HLA-A* cosmids contain highly deteriorated pseudogenes approximately 10 to 15 kb 3′ to their *HLA-A* genes. In CosRS28, the pseudogene resides on a 5.7-kb *Eco*RI fragment separated from *HLA-A2* by fragments of 6.5 and 0.5 kb. Like the *HLA-A3* cosmid, CosRS28 also contains a class I-hybridizing fragment upstream from *HLA-A2*; however, this 12-kb *Eco*RI fragment contains two class I pseudogenes (unpublished data); whether any of the pseudogenes in CosRS28 are expressed as RNA is not known. CosRS6, like the placental lambda clone, also contains a deteriorated pseudogene; this gene is

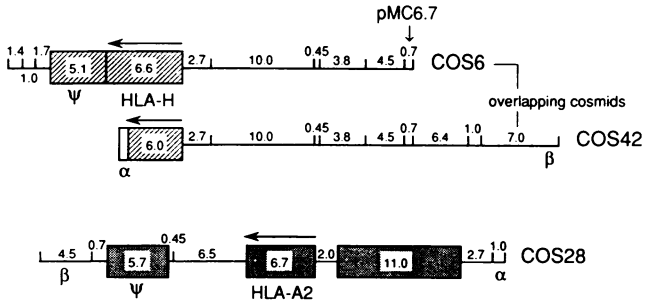

FIG. 2. *Eco*RI restriction maps of the *HLA-H* and *HLA-A2* cosmids. The shaded blocks represent *Eco*RI fragments which hybridized to class I probes. The *HLA-H* gene carried on a 6.6-kb *Eco*RI fragment in CosRS6 but fused to a vector sequence in CosRS42 and the *HLA-A2* gene carried on a 6.7-kb *Eco*RI fragment in CosRS28 are marked. The numbers refer to the sizes of the fragments in kilobases. α and β (in CosRS28 and CosRS42) refer to the 0.4- and 1.8-kb *Eco*RI-*Bam*HI pcos2EMBL vector fragments which became ligated to partially *Sau*3A-digested genomic DNA, thus forming end fusion fragments. CosRS6 was derived from a cosmid library prepared from partially *Eco*RI-digested DNA ligated into the *Eco*RI site of vector pJB8. The horizontal arrows show the transcriptional orientation of the genes. The orientation of *HLA-A2* within the cosmid was determined by digesting the clone with *Kpn*I, which cuts the gene at the end of exon 2, followed by consecutive probing with pHLA-2a.1 and the 2.0-kb *Eco*RI fragment located 5′ of the *HLA-A2* gene. pMC6.7 marks the fragment used as a locus-specific probe on pulsed-field gels. This fragment was subcloned into pUC19 for subsequent propagation.

carried on a 5.1-kb *Eco*RI fragment located immediately 3′ of *HLA-H*. Unlike with *HLA-A2*, we were unable to identify a class I gene within almost 30 kb of DNA 5′ of *HLA-H*.

The 2-kb *Eco*RI fragment immediately 5′ to *HLA-A2* hybridized to the 2.7-kb 5′-flanking fragment from CosRS6, even after high-stringency washing (data not shown). Whether homology extends into the adjacent large *Eco*RI fragments of both cosmids was not determined. In any event, the overall structure of the cosmids, as well as the nucleotide sequences presented above, suggests that these two genes, and perhaps others, arose as the result of either a gene conversion covering more than 6 kb or, more probably, recent duplication of an ancestral gene and its immediately adjacent DNA segments. It is also possible that *HLA-H* is more closely related to either one or both of the pseudogenes carried on the large *Eco*RI fragment of CosRS28 than it is to *HLA-A2*. This notion is supported by the finding that an oligonucleotide probe derived from exon 8 of *HLA-H* (nucleotides 4374 to 4393) consistently hybridized more strongly to the 12-kb *Eco*RI fragment than to the 6.7-kb fragment containing *HLA-A2*.

Since the *HLA-H* and *HLA-A* genes may have arisen by duplication of an ancestral class I gene, we mapped both CosRS6 and CosRS28 by pulsed-field gel electrophoresis. Using the nonrepetitive 0.7-kb *Eco*RI fragment located 5′ of *HLA-H* as a probe (pMC6.7), we observed its hybridization to a 300-kb *Sal*I fragment, a 500-kb *Bss*HII fragment, and a 1,200-kb *Nru*I fragment (Fig. 3A). It also hybridized to a very large *Mlu*I fragment (>1,200 kb) which ran in the compression region at the top of the gel. The 0.7-kb band was also found to hybridize to the 540-kb *Not*I fragment (data not shown) which was earlier found to contain part of the human MHC (44). The remainder of the class I region in 3.1.0 cells is composed of a smaller *Not*I fragment of 190 kb carrying *HLA-E* (44, 68) and a larger 1,090-kb *Not*I fragment

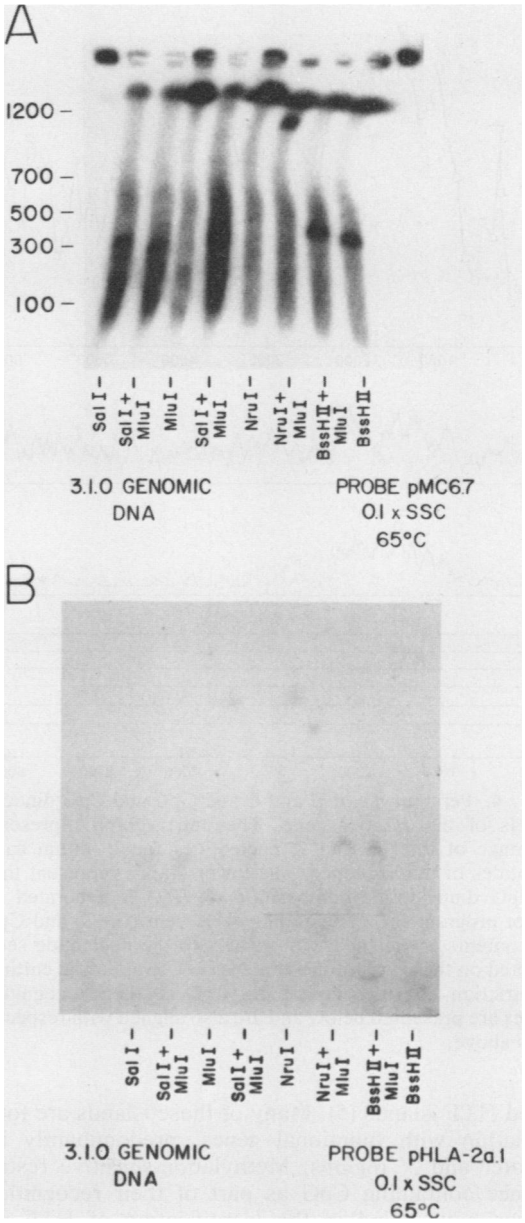

FIG. 3. Pulse-field gels (OFAGE) of digested 3.1.0 genomic DNA blotted and probed with low-copy locus-specific probes. (A) The *HLA-H* locus-specific probe consisted of a 0.7-kb *Eco*RI fragment located approximately 21 kb 5′ of the *HLA-H* gene (Fig. 2). Enzymes used for digestion are marked at the bottom, and molecular sizes in kilobases are to the left. *Saccharomyces cerevisiae* chromosomes and lambda multimers served as size standards. (B) A blot identical to that in panel A, stripped and probed with *HLA-A2*-specific probe pHLA-2a.1.

carrying *HLA-B*, *HLA-C* (44), and the *TNF* genes (G. Gillespie, unpublished data). With the *HLA-A2*-specific probe, identical *Sal*I, *Bss*HII, *Not*I, and *Nru*I fragments were recognized (Fig. 3B). Therefore, both genes are located within 300 kb of each other; however, the exact distance separating them is not known and may be considerably less. The 0.7-kb *Eco*RI probe from CosRS6 also hybridized to DNA from the hybrid mouse-6 line but not to mouse DNA (data not shown).

Regions rich in the unmethylated dinucleotide CpG are

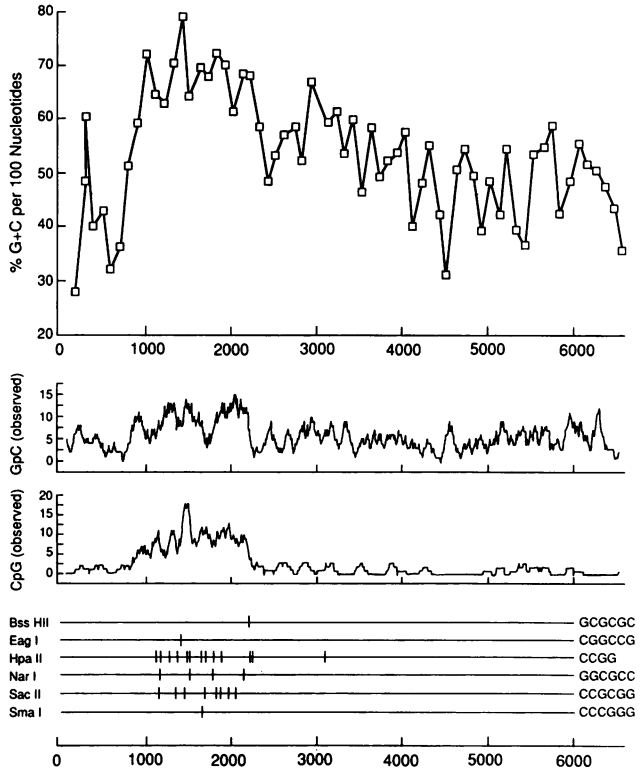

FIG. 4. Percentages of G and C and GpC and CpG dinucleotide contents of the *HLA-H* gene. The upper graph represents the percentage of total G and C nucleotides found within each 100 nucleotides of the sequence. The lower graphs represent the GpC and CpG dinucleotide composition of *HLA-H* generated by the Statplot program (14). The graphs of percent (G+C) and GpC and CpG contents are aligned with respect to the nucleotide sequence presented on the *x* axis of the percent G+C graph. The cutting sites of restriction enzymes containing CpG in their recognition sequences are presented below and are also aligned with respect to the graphs above.

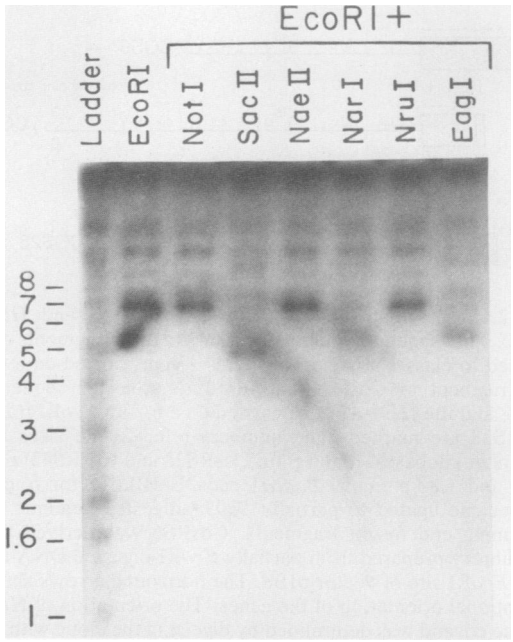

FIG. 5. 3.1.0 genomic DNA restricted with *Eco*RI and *Eco*RI plus rare CpG-recognizing restriction endonucleases. The enzymes used for digestion are marked above the lanes. Molecular sizes (in kilobases) are to the left. The final washing conditions of the filter were 0.2× SSC at 65°C for 2 h. *HLA-H* was used as the probe.

termed HTF islands (5). Many of these islands are found in association with functional genes, predominantly in the promoter and 5′ regions. Methylation-sensitive restriction enzymes containing CpG as part of their recognition sequences have aided in the identification of HTF island-containing functional genes (46, 62), including novel genes within the *H-2K* region of the mouse MHC (1). HTF islands have also recently been associated with expressed (as well as uncharacterized) HLA class I genes (60). Figure 4 shows a plot of the CpG dinucleotide concentration along the *HLA-H* gene. The GpC content roughly parallels the percentage of G and C throughout the gene; however, the CpG content is significantly depleted in the 3′ area compared with the rest of the sequence. The clustering of CpG dinucleotides in the 5′ end of MHC genes (72) and functional α-globin genes (6) has been previously noted. The extent of CpG dinucleotide clustering in *HLA-H* is equivalent to that found within both the *HLA-A2* and *HLA-B27* genes (data not shown).

The CpG dinucleotides are not restricted to the promoter area but are more abundant in the 5′-coding and intron sequences of the gene. The elevated CpG content extends into the beginning of intron 3, where it rapidly decreases. Positions of some of the restriction sites for methylation-sensitive CpG restriction endonucleases are presented below the graph of the *HLA-H* gene in Fig. 4; *Sac*II (CCGCGG) and *Narl* (GGCGCC) sites are particularly abundant. Although *Sac*II sites are expected at a frequency of $2.5/10^5$ nucleotides (based on data present by Lindsay and Bird [46]), the *HLA-H* gene has 8 *Sac*II sites. All of the *Sac*II sites are clustered between positions 1135 (between the CCAAT and TATAA boxes) and 2036 (exon 3) (Fig. 4).

We undertook double digestions of genomic DNA to analyze whether the CpG-rich sites contained within the *HLA-H* gene were methylated. *Eco*RI-*Sac*II double digests of genomic DNA removed the prominent hybridizing band of 6.6 kb after high-stringency washing (Fig. 5; probe-*HLA-H*). The 6.6- to 6.7-kb *Eco*RI fragment includes not only *HLA-H* but also the *HLA-A2* and *HLA-B27* genes. All three genes contain unmethylated *Sac*II sites at approximately equivalent positions. After digestion, a new fragment of about 4.9 kb appeared which represents the 3′ portion of *HLA-H* and probably other digested class I genes, including *HLA-A2* and *HLA-B27*. Double digestion with *Eco*RI-*Nar*I and *Eco*RI-*Eag*I also removed the prominent 6.6- to 6.7-kb band. However, *Bss*HII, which cuts *HLA-H* only once near the end of exon 3, repeatedly failed to digest the *Eco*RI fragment carrying this gene (data not shown).

To determine whether the *HLA-H* gene is transcribed in human cells, we prepared a cDNA library from poly(A)$^+$ RNA of 3.1.0 and examined the class I transcripts from this library. Two cDNA clones were obtained that were derived from *HLA-H*; the sequence of one is compared with that of *HLA-H* in Fig. 6. Both clones started identically within intron 3 of the gene (approximately 100 nucleotides 5′ of exon 4, at position 2691). The splices between exons 4 and 5, 5 and 6, 6 and 7, and 7 and 8 were all at exactly the locations corresponding to those seen in *HLA-A2* cDNA. Both clones terminated approximately 40 nucleotides before the polyadenylation signal, at position 4442. One unexplained feature of the cDNA sequences was that by dideoxy sequencing

```
                5' intron3
                                                                    *
cDNA1 start ggccacatgcgtgctggtggagtgtcccatgacagatgcaaaatgcctgaatttttctgactcttcccgtcag  72
HLA-H       ----------------------------------------------------------------t-----

            exon4
cDNA1       ACCCCCCCCAAGACACATATGACCCACCACCCCATCTCTCTGACCATGAGGCCACCCTGAGGTGCTGGGCCCTGG 144
HLA-H       ----------------------------------------------------------------------
cDNA1       GCTTCTACCCTGCCGGAGATCACACTGACCTGGCAGCGGGATGGGGAGGACCAGACCCACACACGGAGCTCAT 216
HLA-H       ----------------------------------------------------------------------
cDNA1       GGAGACCAGGCCTGCAGGGGATGGAACCTTCCAGAAGTGGGCGGCTGTGGTGGTGCCTTCTGGAGAGGAGCA 288
HLA-H       ----------------------------------------------------------------------
cDNA1       GAGATACACCTGCCATGTGCAGCATGAGGGTCTGCCAGAGCCCCTCACCCTGAGATGGG               347
HLA-H       ------------------------------------------------------------

            exon5
cDNA1       AGCCATCTTCCCAGCCCACCATCCCCATCGTGGGCATCGTTGCTGGCCTGGTTCTACTTGTAGCTGTGGTCA 419
HLA-H       ----------------------------------------------------------------------
cDNA1       CTGGAGCTGTGGTCGCTGCTGTAATGTGGAGGAAGAAGAGCTCAG                            464
HLA-H       ------------------------------------------------

            exon6
cDNA1       ATAGAAAAGGAGGGAGCTACTCTCAGGCTGCAA                                        497
HLA-H       ---------------------------------

            exon7
cDNA1       GCGGCAACAGTGCCCAGGGCTCTGATGTGTCTCTCACGGCGTGAAAG                          544
HLA-H       ------------------------------------------------

            exon8
cDNA1       TGTGAGACAGCTGCCTTGTGTGGGACTGAGAGGCAAGATTTGTTCACACCTTCCCTTTGTGACTTGAAGAAC 616
HLA-H       ----------------------------------------------------------------------
cDNA1       CCTGACTTTCTGCAAAGGCACCTGAATGTGTCTGTGTTCCTGTAGGCATAATGTGTGGAGGAGGGGAGACCA 688
HLA-H       ----------------------------------------------------------------------
cDNA1       ACCCACCCTCATGTCCACCATGACCCTCTTCCCCACGCTGATCTGTGTTCCCTCCCCAATCATCTTTCCTGT 760
HLA-H       ----------------------------------------------------------------------
cDNA1       TCCAGAGAGGCGGGGCTGAGATGTCTCCATCTTTTTCTCAACTTTATGTGCACTGAGCTGTAACTTCTTACT 832
HLA-H       ----------------------------------------------------------------------
cDNA1       TCCCTCTTAAAATTAGAATCTGAGTAAACATTTACTTTTTCAAATTCTTGCCATGAGAGGTTGATGACTT3' 902
                                                                              end
```

FIG. 6. Nucleotide comparison of the 3.1.0 *HLA-H* cDNA clone (cDNA1) and the *HLA-H* gene. The cDNA sequence began within intron 3, at nucleotide position 2691 with respect to the *HLA-H* sequence. All nucleotides within the exons were identical, including the deletions within exons 4 and 7. The only difference, the T-to-C transition within the intron 3 sequence of cDNA1 (position 2757 in *HLA-H*), is marked by an asterisk.

both clones showed a C residue at position 2757, while the corresponding residue in the genomic clones from the placenta, 3.1.0 (CosRS6 and CosRS42), and JY (16) libraries was a T. Although this demonstrates that the gene is transcribed, albeit in an unconventional fashion, neither of these cDNA clones could encode a class I-like gene product.

To further investigate transcription of *HLA-H*, we analyzed RNA transcripts in bulk cultures of mouse L cells transfected with pHLA-H, pB27, and hybrid constructs composed of both pHLA-H and pB27 (Fig. 7). Northern blots of RNAs from pHLA-H transfectants probed with an *HLA-H* genomic probe showed the presence of some full-length transcripts (Fig. 8A). RNase protection studies with a

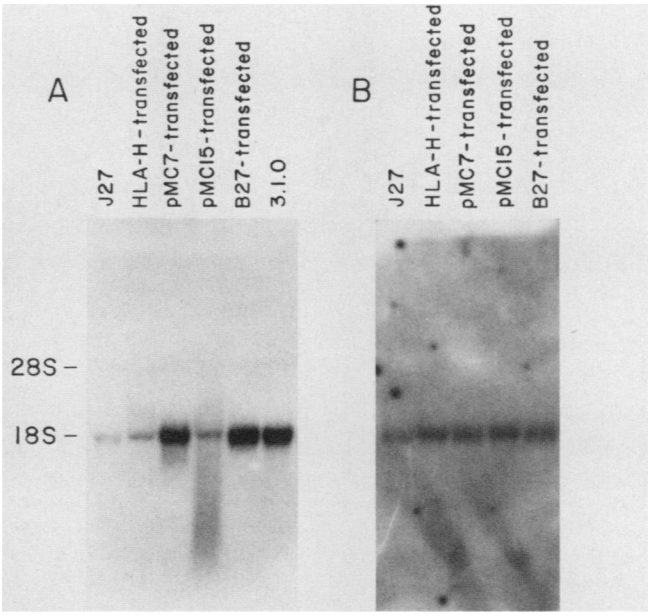

FIG. 8. Northern blot of total cytoplasmic RNA from L cells (J27) transfected with pHLA-H, pB27, and the hybrid clones pMC7 and pMC15. (A) The Northern blot probed with the labeled *HLA-H* genomic fragment. Washing was with 2× SSC at room temperature for 30 min, followed by 0.1× SSC at 65°C for 10 min. (B) The same Northern blot as in panel A, stripped and reprobed with *H-2K*^{k} cDNA. Washing was as for panel A. J27 and each transfectant are marked above the lanes. 3.1.0 total cytoplasmic RNA in panel A served as a positive control.

probe spanning exon 1 (73 nucleotides long) and continuing 5′ of the gene (Fig. 9) showed that about half of the transcripts initiated at the site predicted by analogy to full-length *H-2K*^{d} (41) and *HLA-A2* cDNA clones (M.J.C., unpublished data) (Fig. 10). This position is approximately 25 nucleotides upstream of the first ATG of exon 1; the size of the expected protected fragment is therefore 73 + 25 = 98

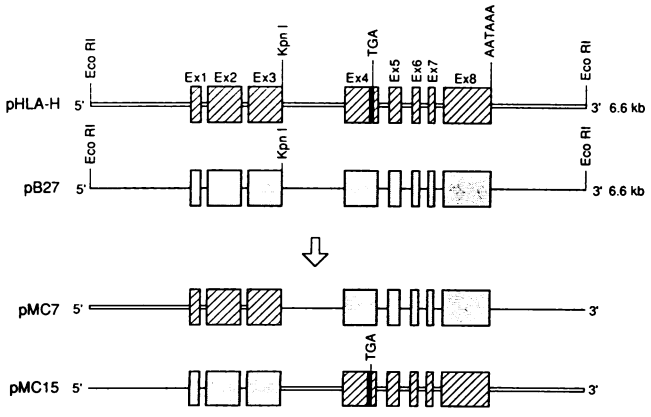

FIG. 7. Construction of hybrid *HLA-H* genes. The boxes represent exons which are labeled at the top. The termination codon in exon (Ex) 4 of *HLA-H* is represented as a solid bar. The single *Kpn*I site at the end of exon 3 in both pHLA-H and pB27 is shown. The hybrid clones pMC7, containing the 5′ *Eco*RI-*Kpn*I fragment of pHLA-H joined to the 3′ *Kpn*I-*Eco*RI fragment of pB27, and pMC15, containing the 5′ *Eco*RI-*Kpn*I fragment of pB27 joined to the 3′ *Kpn*I-*Eco*RI fragment of pHLA-H, are shown.

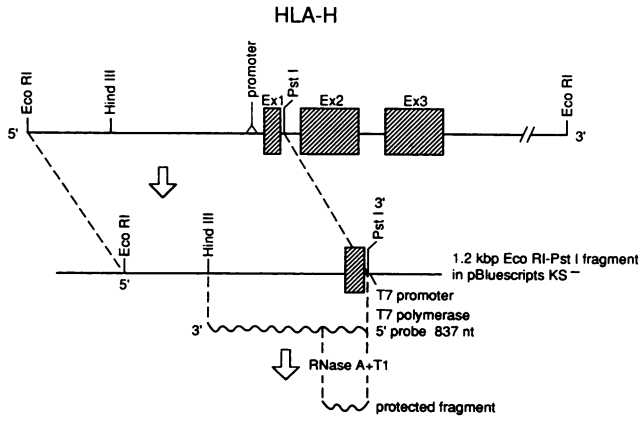

FIG. 9. Schematic representation of the RNase protection assay. The 5′ *Eco*RI-*Pst*I fragment (1.2 kilobase pairs [kbp]) was subcloned into pBluescript KS(−) and then linearized with *Hin*dIII. T7 RNA polymerase was used to synthesize transcripts of 838 nucleotides (nt) containing exon (Ex) 1 (73 nucleotides) plus the 5′ promoter and flanking sequences. The size of the expected protected fragment was approximately 100 nucleotides (73 plus approximately 25 nucleotides 5′ of the first ATG in exon 1; see text). kbp, Kilobase pair.

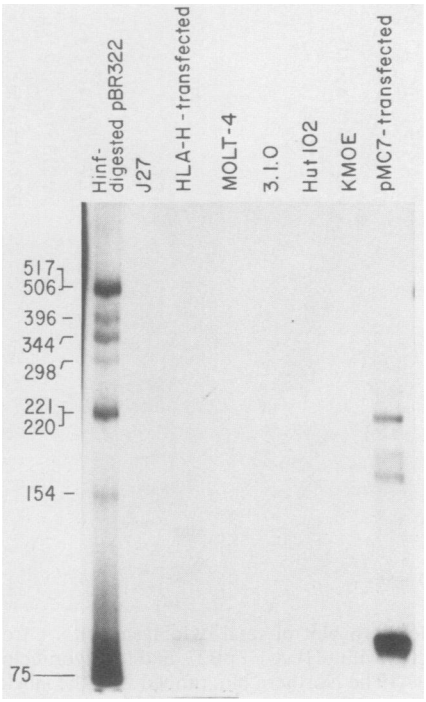

FIG. 10. RNase protection assay. Cell lines and transfectants are marked above the lanes. End-filled *Hin*fI-digested pBR322 served as size markers. The approximate sizes of the major protected fragments in the pHLA-H and pMC7 lanes were 90 to 100 nucleotides. No bands were observed below this fragment on other gel runs. MOLT-4, Hut102, and KMOE are human cell lines. bp, Base pairs.

nucleotides. The other half of the transcripts initiated at a somewhat upstream site (Fig. 10). The amount of the transcripts accumulated was no more than 5% of the amount in L cells transfected with the *HLA-B27* gene alone.

Composite genes were constructed by cutting the *HLA-H* and *HLA-B27* genes at their unique and homologous *Kpn*I sites at the exon 3-intron 3 junction and joining the 5′ end of each gene to the 3′ end of the other (Fig. 7). Cells transfected with the composite gene with its 5′ portion, including the promoter derived from *HLA-H*(pMC7), accumulated as much RNA as did cells transfected with the functional *HLA-B27* gene (Fig. 8A). Most of the RNA transcribed from this hybrid gene initiated at the expected downstream site (Fig. 10); however, several other minor bands of protected RNA were also observed. In contrast, the cells transfected with the composite gene derived from the 5′ half of the *HLA-B27* gene and the 3′ half of the *HLA-H* gene accumulated no more full-length RNA than did cells transfected with the native *HLA-H* gene (Fig. 8A), suggesting that the defect in RNA accumulation in these cells was due not to promoter sequences but rather to sequences downstream of exon 3. It may be noted that in the pMC15 lane, the *HLA-H* probe hybridized to a considerable smear of degraded RNA. No such smear was evident when *H-2K*^{k} cDNA was used as a probe on the same blot (Fig. 8B). Ethidium bromide staining of the gel before blotting showed that the RNA was intact. This suggests that the degradation was limited to the transcripts derived from the hybrid gene-containing plasmid, pMC15.

DISCUSSION

The *HLA-H* gene resembles previously sequenced class I genes, except for a single deleted nucleotide in exon 4 that creates a garbled amino acid sequence for much of exon 4 and an in-frame termination at the end of this exon. Premature termination codons are characteristic of mouse *Qa* class I genes (38, 63, 70), as well as at least one recently described human gene (22). In the human gene, however, the termination was found in exon 6 instead of exon 5, as in mouse *Qa* genes.

Unmethylated CpG sequences that make up HTF islands are found in association with many expressed genes (5). Most genomic fragments that carry class I genes are cleaved by methylation-sensitive CpG enzymes (reference 60; this report). The indication that most class I genes that contain CpG-rich islands are transcribed is supported by identification of specific transcripts on Northern blots or isolation of specific cDNA clones. In our own laboratory, we have isolated cDNA clones for *HLA-E* (source, B-lymphoblastoid line JY; H. Shukla, A. Swaroop, and S. Weissman, unpublished data), *HLA-6.0* (source, fetal eye library; A. Swaroop, H. Shukla, R. Srivastava, and S. Weissman, unpublished data), and *HLA-H*.

The *HLA-H* gene gives rise to low levels of transcripts in transfected L cells and the human lymphoblastoid line 3.1.0. The fact that the 6.6-kb *Eco*RI genomic fragment carrying the gene was cleaved by *Sac*II and *Nar*I indicates that the *HLA-H* gene contains an HTF island. The *Bss*HII site beginning at position 2172 was not cut in repeated experiments and may lie beyond the 3′ end of the island.

The characterized cDNA clones of *HLA-H* begin at a position within intron 3 approximately 100 nucleotides upstream from exon 4 and are derived from a transcript, because subsequent introns are spliced out in the pattern predicted by mRNA for conventional class I genes. The cDNA clones have a C at position 2755, while the *HLA-H* gene has a T; the rest of the compared sequences are identical. The simplest explanation is that the single-base change is the result of a reverse transcriptase error or a cloning artifact.

There are abundant reports that alternate splicing occurs within both mouse and human class I gene transcripts (37, 40). Full-length cDNA clones containing either intact intron 1 or 2 have been reported for mouse *TL* class I genes (10, 51). *TL* transcripts containing unspliced intron 3 sequences have also been observed for transfected *T3*^{c} and *T13*^{c} *TL* genes (M. J. Chorney, H. Mashimo, Y. Bushkin, and S. G. Nathenson, J. Immunol., in press), as well as for the *T3*^{b} gene expressed in a leukemia line (H. Mashimo, Ph.D. thesis, Albert Einstein College of Medicine, New York, N.Y., 1988).

The *HLA-H* cDNA clone may have derived from shortened transcripts which initiated within intron 3 of the *HLA-H* gene. In support of this hypothesis is the finding that a 5′-truncated *HLA-B7* gene directed the synthesis of shortened transcripts after transfection into L cells (78). Ganguly et al. (21) have identified enhancer elements located in both the 5′ promoter region and the 5′ and 3′ ends of intron 3 of the *HLA-B7* gene. In addition to enhancer-binding sites, secondary RNA polymerase II transcription initiation sites may also reside within this region. If this is true of most class I genes, including *HLA-H*, then the CpG-rich islands would be flanked by the natural promoter at the 5′ end (containing enhancer A, IRS, TATA, and CCAAT sequences) and enhancer elements located 3′ within intron 3. Bird has

suggested that CpG-rich DNA-binding factors could competitively inhibit the binding of methylases to these regions, thereby limiting CpG depletion by methylation and subsequent deamination (5). An alternative speculation is that the promoter-binding factors themselves could potentially hinder the entrance of methylases and A-T-prone repair enzymes to the CpG-rich HTF islands located within the 5′ ends of the genes.

HLA-H gives rise to relatively small transcript amounts in transfected L cells. The experiments involving transfection of hybrid genes suggest that the promoter of *HLA-H* is potentially as effective in these cells as is the promoter of the functional *HLA-B27* gene, despite the deviations in the promoter sequence from conventional class I genes that were noted above. An effect of sequences downstream of the initiation site on the level of expression of the antigen encoded by a composite murine *Qa* region gene has been observed, but the mechanisms of that effect are not entirely clear (71). Possible explanations for the effect of downstream sequences of *HLA-H* in diminishing the amount of transcript would be a defective downstream enhancer or a polarity effect due to a translation termination codon in an interior exon. The latter is an attractive hypothesis, particularly since downstream enhancers have not been described for class I genes and premature translation termination has been shown to be responsible for the instability of triosephosphate isomerase (13) and β-globin transcripts (50). A number of nonclassical class I genes of mice and humans have termination codons within internal exons, and an effect of these termination codons in down modulating the levels of gene transcripts in cells could be an important mechanism in limiting the expression of these genes.

The sequence of *HLA-H* suggests that the gene is more closely related to *HLA-A2* than to any other known class I gene. If one assumes that the evolutionary rate for silent substitutions (V_s) is 5.5×10^{-9} per site per year (24), the time of divergence can be calculated by $T = K_s/2V_s$, where K_s is the average number of nucleotide differences per potential site corrected for multiple substitutions. This equation produces a time of divergence of approximately 8.8 million years. This time precedes human-chimpanzee separation, which is estimated to have occurred approximately 5×10^6 to 7×10^6 million years ago, but is concomitant with the separation of many of the *HLA-A* and *HLA-B* alleles (42, 52). Because *HLA-H* has retained a functional promoter and an abundance of CpG dinucleotides in its 5′ end similar to those of functional HLA class I genes, the deletion in exon 4 may have been a very recent event. In consideration of the *trans*-species theory of gene diversification (30, 52), *HLA-H* may be a pseudogene in humans but may still be functional in other primate species and could thereby be a member of the *HLA-A*-like allelic family.

It has been suggested that in mice, *H-2K* and *Qa* genes are closely related, since flanking probes from *H-2K* clones hybridize to *Qa* class I cosmids (74). In humans, we have observed that both the 0.7-kb *Eco*RI fragment 3′ and the 2.0-kb *Eco*RI fragment 5′ of *HLA-A2* hybridized after high-stringency washing to several of our cosmids containing divergent class I genes (M.J.C., unpublished data). One such clone that hybridized to both probes is the cosmid containing the *HLA-6.0* gene. This suggests that, like the *H-2K* gene in mice, ancestral *HLA-A* genes may have been actively involved in duplication events, giving rise to a putative human *Qa* gene cluster composed in part of *HLA-6.0* and *HLA-H*. CosRS6 did not hybridize to the 0.7-kb probe, although it hybridized to the 2.0-kb *Eco*RI fragment. On the basis of the maps, one can envision deletions postduplication occurring independently in the *HLA-H* cosmid: one between the 6.6-kb *Eco*RI fragment and the 5.1-kb fragment carrying the deteriorated pseudogene and the other 3′ of the pseudogene fragment itself.

In the 3.1.0 genome, low-copy probes from both the *HLA-A2* and *HLA-H* cosmids hybridized to identical large DNA fragments. The result of these experiments show that the *HLA-A2* and *HLA-H* genes reside on common *Sal*I, *Bss*HII, and *Not*I fragments and lie within at least 300 kb of each other.

The question arises as to whether the gene directs the synthesis of a protein product. Because of premature termination, the protein, if expressed, would be approximately 30 kilodaltons in size. It is unlikely that the HLA-H protein would interact directly with the lipid bilayer; instead, it would most probably be secreted. Reports of HLA class I molecules in serum smaller in size than the typical 40-kilodalton transmembrane proteins exist (15, 59); some of these molecules may, however, arise as a result of deletion of the exon 5 sequence (encoding the hydrophobic transmembrane domain) from otherwise normal class I transcripts (37).

Some mouse *Qa* genes which contain premature termination codons are linked to the cell surface via a phosphatidylinositol linkage. In $Q7^b$, attachment is mediated by the carboxyl group of an aspartic acid residue located at position 295 (73). *HLA-H*, if expressed, may possibly be linked by a similar mechanism. After the frameshift in exon 4, the rest of the sequence contains one aspartic and two glutamic acid residues (Fig. 1, legend), any of which could supply a carboxyl group for attachment to the membrane surface.

Like the translated sequence of *HLA-12.4*, that of *HLA-H* also predicts a phenylalanine at position 162 in place of a conserved cysteine. As pointed out originally by Malissen et al. (47), this substitution would result in destruction of the alpha 2 domain intrachain disulfide bond found in all class I proteins. As Klein and Figueroa have argued with respect to mouse *Qa* and *T1a* genes (31), this novel human gene may indeed be nothing more than a pseudogene which has fortuitously retained a functional promoter but awaits ultimate extinction from the human genome. Nonetheless, the mechanism by which the abundance of stable transcripts of *HLA-H* is decreased is of general interest. *HLA-H* provides a favorable system for further analysis of this phenomenon, since a marked effect can be seen in transfected cells.

ACKNOWLEDGMENTS

We thank Barbara Gramenos and Leif Madsen for excellent technical assistance and expert preparation of the manuscript, H. Vasavada and S. Ganguly for helpful discussions and comments, and R. Srivastava for work in isolating the cosmid clones.

This work was supported by Public Health Service grant 5 R35 CA42556-03 from the National Cancer Institute. Michael J. Chorney was the recipient of Public Health Service postdoctoral fellowship GM09964-03 from the National Institutes of Health.

LITERATURE CITED

1. **Abe, K., J.-F. Wei, F.-S. Wei, Y.-C. Hsu, H. Uehara, K. Artzt, and D. Bennett.** 1988. Searching for coding sequences in the mammalian genome: the *H-2K* region of the mouse MHC is replete with genes expressed in the embryo. EMBO J. **7**: 3441–3449.
2. **Anonymous.** 1987. Nomenclature for factors of the HLA system, 1987. Immunogenetics **28**:391–309.
3. **Baldwin, A. S., Jr., and P. A. Sharp.** 1987. Binding of a nuclear factor to a regulatory sequence in the promoter of the mouse

H-2K^b class I major histocompatibility gene. Mol. Cell. Biol. **7**:305–313.
4. **Berger, S. L., and A. R. Kimmel (ed.)** 1987. Guide to molecular cloning. Methods Enzymol. **152**:307–389.
5. **Bird, A. P.** 1986. CpG-rich islands and the function of DNA methylation. Nature (London) **321**:209–213.
6. **Bird, A. P., M. H. Taggart, R. D. Nicholls, and D. R. Higgs.** 1987. Non-methylated CpG-rich islands at the human α-globin locus: implications for evolution of the α-globin pseudogene. EMBO J. **6**:999–1004.
7. **Biro, A., P. Pan, A. Sood, R. Kole, V. B. Reddy, and S. M. Weissman.** 1983. Sequences of the human repetitive DNA, non-α globin genes, and major histocompatibility locus genes. III. The major histocompatibility complex. Cold Spring Harbor Symp. Quant. Biol. **47**:1082–1086.
8. **Carle, G. F., M. Frank, and M. V. Olson.** 1986. Electrophoretic separations of large DNA molecules by periodic inversion of the electric field. Science **232**:65–68.
9. **Carroll, M. C., P. Katzman, E. M. Alicot, B. H. Koller, D. E. Geraghty, H. T. Orr, J. L. Strominger, and T. Spies.** 1987. Linkage map of the human major histocompatibility complex including the tumor necrosis factor genes. Proc. Natl. Acad. Sci. USA **84**:8535–8539.
10. **Chen, Y.-T., Y. Obata, E. Stockert, and L. J. Old.** 1985. Thymus-leukemia (TL) antigens of the mouse. Analysis of TL mRNA and TL cDNA from TL$^+$ and TL$^-$ strains. J. Exp. Med. **162**:1134–1148.
11. **Church, G. M., and W. Gilbert.** 1984. Genomic sequencing. Proc. Natl. Acad. Sci. USA **81**:1991–1995.
12. **Coppin, H. L., D. W. Denny, S. M. Weissman, and H. O. McDevitt.** 1985. *HLA-B* locus polymorphism: studies with a specific hybridization probe. Proc. Natl. Acad. Sci. USA **82**: 8614–8618.
13. **Daar, I. O., and L. E. Maquat.** 1988. Premature translation termination mediates triosephosphate isomerase mRNA degradation. Mol. Cell. Biol. **8**:802–813.
14. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the Vax. Nucleic Acids Res. **12**:387–395.
15. **Dobbe, L. M. E., N. J. Stam, J. J. Neefjes, and M. J. Giphart.** 1988. Biochemical complexity of serum HLA class I molecules. Immunogenetics **27**:203–210.
16. **Duceman, B., and A.-M. Wang.** 1988. A novel human major histocompatibility complex class I-type DNA sequence. Nucleic Acids Res. **16**:10391.
17. **Ellis, S. A., C. Taylor, and A. McMichael.** 1984. Recognition of HLA-B27 and related antigen by a monoclonal antibody. Hum. Immunol. **5**:49–59.
18. **Favaloro, T., R. Treisman, and R. Kamen.** 1980. Transcription maps of polyoma virus-specific RNA: analysis by two dimensional nuclease S1 gel mapping. Methods Enzymol. **65**:718–749.
19. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**:6–13.
20. **Flavell, R. A., H. Allen, B. Huber, C. Wake, and G. Widera.** 1985. Organization and expression of the MHC of the C57 black/10 mouse. Immunol. Rev. **84**:29–50.
21. **Ganguly, S., H. A. Vasavada, and S. M. Weissman.** 1989. Multiple enhancer-like sequences in the *HLA-B7* gene. Proc. Natl. Acad. Sci. USA **86**:5247–5251.
22. **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**:9145–9149.
23. **Hanahan, D., and M. Meselson.** 1980. Plasmid screening at high colony density. Gene **10**:63–67.
24. **Hayashida, H., and T. Miyata.** 1983. Unusual evolutionary conservation and frequent DNA segment exchange in class I genes of the major histocompatibility complex. Proc. Natl. Acad. Sci. USA **80**:2671–2675.
25. **Hood, L., M. Steinmetz, and B. Malissen.** 1983. Genes of the major histocompatibility complex. Annu. Rev. Immunol. **1**: 529–568.
26. **Johnson, D. A., J. W. Gautsch, J. R. Sportsman, and J. H. Elder.** 1984. Improved technique utilizing nonfat dry milk for analysis of proteins and nucleic acids transferred to nitrocellulose. Gene Anal. Tech. **1**:3–8.
27. **Jordan, B. R., D. Caillol, M. Damotte, T. Delovitch, P. Ferrier, B. Kahnperles, F. Kourilsky, C. Layet, P. Le Bouteiller, F. A. Lemonnier, M. Malissen, C. N'Guyen, J. Sire, R. Sodoyer, T. Strachan, and J. Trucy.** 1985. HLA class I genes: from structure to expression, serology and function. Immunol. Rev. **84**:73–92.
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**:261–272.
29. **Klein, J.** 1986. Natural history of the major histocompatibility complex. John Wiley & Sons, Inc., New York.
30. **Klein, J.** 1987. Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. Hum. Immunol. **19**:155–162.
31. **Klein, J., and F. Figueroa.** 1986. Evolution of the major histocompatibility complex. Crit. Rev. Immunol. **6**:295–386.
32. **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**:1–10.
33. **Koller, B. H., D. E. Geraghty, R. Demars, L. Duvick, S. S. Rich, and H. T. Orr.** 1989. Chromosomal organization of the human major histocompatibility complex class I gene family. J. Exp. Med. **169**:469–480.
34. **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 T lymphocytes. J. Immunol. **141**:897–904.
35. **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**:2727–2733.
36. **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**:5175–5178.
37. **Krangel, M. S.** 1986. Secretion of HLA-A and -B antigens via an alternative RNA splicing pathway. J. Exp. Med. **163**:1173–1190.
38. **Kress, M., D. Cosman, G. Khoury, and G. Jay.** 1983. Secretion of a transplantation-related antigen. Cell **34**:189–195.
39. **Kress, M., W.-Y. Liu, E. Jay, G. Khoury, and G. Jay.** 1983. Comparison of class I (H-2) gene sequences. J. Biol. Chem. **258**:13929–13936.
40. **Lalanne, J.-L., M. Cochet, A.-M. Kummer, G. Gachelin, and P. Kourilsky.** 1983. Different exon-intron organization at the 5′ part of a mouse class I gene is used to generate a novel H-2K^k-related mRNA. Proc. Natl. Acad. Sci. USA **80**:7561–7565.
41. **Lalanne, J.-L., C. Delarbre, G. Gachelin, and P. Kourilsky.** 1983. A cDNA clone containing the entire coding sequence of a mouse H-2K^d histocompatibility antigen. Nucleic Acids Res. **11**:1567–1577.
42. **Lawlor, D. A., F. E. Ward, P. D. Ennis, A. P. Jackson, and P. Parham.** 1988. *HLA-A* and *B* polymorphisms predate the divergence of humans and chimpanzees. Nature (London) **335**: 268–271.
43. **Lawrance, S., H. K. Das, J. Pan, and S. M. Weissman.** 1985. The genomic organization and nucleotide sequence of the HLA-SB(DP) alpha gene. Nucleic Acids Res. **13**:7515–7527.
44. **Lawrance, S. K., C. L. Smith, R. Srivastava, C. R. Cantor, and S. M. Weissman.** 1987. Megabase-scale mapping of the HLA gene complex by pulsed field gel electrophoresis. Science **235**: 1387–1390.
45. **Levine, F., H. Erlich, B. Mach, R. Leach, R. White, and D. Pious.** 1985. Deletion mapping of HLA and chromosome 6p genes. Proc. Natl. Acad. Sci. USA **82**:3741–3745.
46. **Lindsay, S., and A. P. Bird.** 1987. Use of restriction enzymes to detect potential gene sequences in mammalian DNA. Nature (London) **327**:336–338.
47. **Malissen, M., B. Malissen, and B. R. Jordan.** 1982. Exon/intron organization and complete nucleotide sequence of an HLA gene. Proc. Natl. Acad. Sci. USA **79**:893–897.
48. **Maniatis, F., R. C. Hardison, E. Lacy, J. Lauer, C. O'Connel, and D. Quon.** 1978. The isolation of structural genes from

libraries of eucaryotic DNA. Cell **15**:687–701.

49. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual, p. 213–246. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
50. **Maquat, L. E., A. J. Kinniburgh, E. A. Rachmilewitz, and J. Ross.** 1981. Unstable beta-globin mRNA in mRNA-deficient β^0-thalassemia. Cell **27**:543–550.
51. **Matsuura, A., F.-W. Shen, D. A. Fisher, L. Hood, and E. A. Boyse.** 1987. Transcripts of *Tla* genes. Immunogenetics **25**: 411–415.
52. **Mayer, W. E., M. Jonker, D. Klein, P. Ivanyi, G. van Seventer, and J. Klein.** 1988. Nucleotide sequences of chimpanzee MHC class I alleles: evidence for trans-species mode of evolution. EMBO J. **7**:2765–2774.
53. **McMichael, A. J., P. Parham, N. Rust, and F. Brodsky.** 1980. A monoclonal antibody that recognizes an antigenic determinant shared by HLA A2 and B17. Hum. Immun. **1**:121–129.
54. **Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green.** 1984. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12**:7035–7056.
55. **Mizuno, S., J. A. Trapani, B. H. Koller, B. Dupont, and S. Y. Yang.** 1988. Isolation and nucleotide sequence of a cDNA clone encoding a novel HLA class I gene. J. Immunol. **140**:4024–4030.
56. **N'Guyen, C., R. Sodoyer, J. Trucy, T. Strachan, and B. R. Jordan.** 1985. The HLA-AW24 gene: sequence, surroundings and comparison with the HLA-A2 and HLA-A3 genes. Immunogenetics **21**:479–489.
57. **Norrander, Y., T. Kempe, and Y. Messing.** 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene **26**:101–106.
58. **Orr, H. T., and R. DeMars.** 1983. Class I-like HLA genes map telomeric to the HLA-A2 locus in human cells. Nature (London) **302**:534–536.
59. **Pellegrino, M. A., C. Russo, and J. Allison.** 1984. HLA antigens in serum. Methods Enzymol. **108**:614–624.
60. **Pontarotti, P., G. Chimini, C. Nguyen, J. Boretto, and B. R. Jordan.** 1988. CpG islands and HTF islands in the HLA class I region: investigation of the methylation status of class I genes leads to precise physical mapping of the HLA-B and -C genes. Nucleic Acids Res. **16**:6767–6778.
61. **Ragousis, J., K. Bloemer, H. Pohla, G. Messer, E. H. Weiss, and A. Ziegler.** 1989. A physical map including a new class I gene (cda12) of the human major histocompatibility complex (A2/B13 haplotype) derived from a monosomy 6 mutant cell line. Genomics **4**:301–308.
62. **Rappold, G. A., L. Stubbs, S. Labeit, R. B. Crkvenjakov, and H. Lehrach.** 1987. Identification of a testis-specific gene from the mouse t-complex next to a CpG-rich island. EMBO J. **6**: 1975–1980.
63. **Robinson, P. J., D. Bevec, A. L. Mellor, and E. H. Weiss.** 1988. Sequence of the mouse Q4 class I gene and characterization of the gene product. Immunogenetics **27**:79–86.
64. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
65. **Smith, C. L., S. K. Lawrance, G. A. Gillespie, C. R. Castor, S. M. Weissman, and F. C. Collins.** 1987. Strategies for mapping and cloning macroregions of mammalian genomes. Methods Enzymol. **151**:461–489.
66. **Sood, A. K., J. Pan, P. A. Biro, D. Pereira, V. B. Reddy, H. K. Das, and S. M. Weissman.** 1984. Cloning and structure analysis of histocompatibility class I and class II genes, p. 87–108. *In* A. Kumar (ed.), Eukaryotic gene expression. Plenum Publishing Corp., New York.
67. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
68. **Srivastava, R., M. J. Chorney, S. K. Lawrance, J. Pan, Z. Smith, C. L. Smith, and S. M. Weissman.** 1987. Structure, expression, and molecular mapping of a divergent member of the class I HLA gene family. Proc. Natl. Acad. Sci. USA **84**:4224–4228.
69. **Srivastava, R., B. W. Duceman, P. A. Biro, A. S. Sood, and S. M. Weissman.** 1985. Molecular organization of the class I genes of human major histocompatibility complex. Immunol. Rev. **84**:93–121.
70. **Steinmetz, M., K. W. Moore, J. G. Frelinger, B. T. Sher, F.-W. Shen, E. A. Boyse, and L. Hood.** 1981. A pseudogene homologous to mouse transplantation antigens: transplantation antigens are encoded by eight exons that correlate with protein domains. Cell **25**:683–692.
71. **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 antigen. J. Exp. Med. **161**:935–952.
72. **Tykocinski, M. L., and E. E. Max.** 1984. CG dinucleotide clusters in MHC genes and in 5′ demethylated genes. Nucleic Acids Res. **12**:4385–4396.
73. **Waneck, G. L., M. E. Stein, and R. A. Flavell.** 1988. Conversion of a PI-anchored protein to an integral membrane protein by a single amino acid mutation. Science **241**:697–699.
74. **Weiss, E.** 1987. Molecular biology of the mouse Q region. Immunol. Res. **6**:179–191.
75. **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 (London) **310**:650–654.
76. **Wigler, M., R. Sweet, G. K. Sim, B. Wold, A. Pellicer, E. Lacy, T. Maniatis, S. Silverstein, and R. Axel.** 1979. Transformation of mammalian cells with genes for procaryotes and eucaryotes. Cell **16**:777–785.
77. **Winoto, A., M. Steinmetz, and L. Hood.** 1983. Genetic mapping in the major histocompatibility complex by restriction enzyme site polymorphisms: most mouse class I genes map to the Tla complex. Proc. Natl. Acad. Sci. USA **80**:3425–3429.
78. **Yoshie, O., H. Schmidt, P. Lengyel, E. S. P. Reddy, W. R. Morgan, and S. M. Weissman.** 1984. Transcripts of human HLA gene fragments lacking the 5′-terminal region in transfected mouse cells. Proc. Natl. Acad. Sci. USA **81**:649–653.