# Structure, expression, and molecular mapping of a divergent member of the class I *HLA* gene family

(major histocompatibility complex/Qa and TL/pulsed-field electrophoresis/transfection)

Rakesh Srivastava<sup>\*†</sup>, Michael J. Chorney<sup>\*</sup>, Simon K. Lawrance<sup>\*†</sup>, Julian Pan<sup>\*</sup>, Zaid Smith<sup>‡</sup>, Cassandra L. Smith<sup>§</sup>, and Sherman M. Weissman<sup>\*</sup>

Departments of \*Human Genetics and <sup>‡</sup>Cell Biology, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06510; and <sup>§</sup>Department of Human Genetics and Development, Columbia University, 701 West 168th Street, New York, NY 10032

Contributed by Sherman M. Weissman, February 24, 1987

ABSTRACT A class I gene distinct from HLA-A, -B, or -C was identified in a cosmid clone and transfected into mouse L cells. The gene, placed adjacent to the polyoma enhancer, produced a full-length class I mRNA and high levels of a 43-kDa protein in the cytoplasm. The surface expression of the gene product required its association with human  $\beta_2$ -microglobulin. The protein was recognized by a xenoantiserum raised against a mixture of human B- and T-cell lines. The product was also serologically reactive with the HLA framework monoclonal antibodies. The complete nucleotide sequence of the gene was determined and a specific oligonucleotide probe was synthesized. This probe was used to identify a full-length mRNA transcript in a B-lymphoblastoid cell line (JY). The gene was mapped within a 190-kilobase Not I restriction fragment located in the telomeric portion of the human major histocompatibility complex. Distinct features of the gene include (i) the structure of the promoter, (ii) the position of the translation initiation site, (iii) a frameshift mutation at the carboxyl terminus, (iv) the insertion of an Alu repeat element in the eighth exon, (v) divergence in the derived amino acid sequence, and (vi) the lack of expression of the gene in some cells.

The HLA class I region of the human major histocompatibility complex is composed of a family of closely related genes located on the short arm of chromosome 6 (6p21). Although >20 distinct genes or pseudogenes per haploid genome have been identified in this multigene family by molecular cloning (1), the encoded products of only three (HLA-A, -B, and -C) have been detected serologically. These proteins serve as restriction elements in the lysis of virally infected cells by cytotoxic T lymphocytes and are the major determinants in directing tissue graft rejections.

Some of the class I genes map telomeric to the *HLA-A* locus (2), a region that in the mouse contains Qa/TL genes (3). The nucleotide sequence of two, the *HLA12.4* (4) and *LN11* (5), class I pseudogenes was >80% homologous to the *HLA-A*, -*B*, and -*C* genes.

Here we report the cloning, sequencing, molecular mapping, and expression of a human class I gene, the derived amino acid sequence of which revealed a level of divergence from *HLA-A*, -*B*, and -*C* locus proteins comparable to that noted between *H-2* and some of the Qa/TL region genes in the mouse (6).

## MATERIALS AND METHODS

Cell Lines. The human B-lymphoblastoid cell lines 3.1.0 (7), JY (8), and GM3104 (National Institute of General Medical Sciences human cell repository), the T-cell lines

MOLT-4, HuT 78, HuT 102 (all from American Type Culture Collection), and SU (9) and the mouse cell lines LMTK<sup>-</sup> (ATCC) and J27 (10) were maintained in Dulbecco's modified Eagle's medium supplemented with 10% heat-inactivated fetal bovine serum.

The procedures for megabase mapping of the human major histocompatibility complex have been published elsewhere (11). Genomic DNA was electrophoresed along with yeast chromosomes and multimeric forms of phage  $\lambda$  genome as markers. The gel was blotted onto nitrocellulose paper by a standard technique (12).

The preparation of the genomic cosmid libraries from the HLA hemizygous cell line 3.1.0 (HLA-A2, B27, Cw1, DR1, DQwl) is described elsewhere (13). The cosmid RS5 was isolated using as a probe a 6.4-kilobase (kb) EcoRI fragment containing the pseudogene LN11 (5) cloned in  $\lambda$ gt WES. The clone RS5 has an insert of ≈38 kb. An 18-kb EcoRI fragment contained the entire class I sequence. Subclones were made from this clone as follows: an 18-kb EcoRI fragment containing the gene cloned either in pBR322 (5.1) or in the vector poly-ori-Bam [py 5.1; the vector poly-ori-Bam has a 433base-pair (bp) fragment containing the polyoma virus origin of replication and enhancer cloned in the BamHI site of the plasmid pBR322] and a 700-bp terminal EcoRI fragment of the cosmid clone RS5 cloned in pBR322 (5.10). Two additional subclones, 5.1A and 5.1B (see Fig. 2), were made from the 18-kb EcoRI fragment after BamHI digestion and contain the 5' and 3' halves of the class I gene, respectively.

Nucleotide sequencing of subclones 5.1A and 5.1B in the *Sma* I-cut M13 vector M13mp8 was carried out using Sanger's dideoxy method (14).

Two mouse cell lines, LMTK<sup>-</sup> and the LMTK<sup>-</sup>-derived J27 were used in transfection studies (15). The cell line J27 was obtained by Kavathas and Herzenberg (10) after the transfection of the human genomic DNA into LMTK<sup>-</sup> cells and selection for the expression of human  $\beta_2$ -microglobulin. About 10  $\mu$ g of DNA (py5.1; see Fig. 2b) was cotransfected with 50 ng of either herpes simplex virus thymidine kinase gene containing plasmid DNA (in the case of LMTK<sup>-</sup> cells) or pSV2Neo DNA (in the case of J27 cells). Transfectants were maintained either as single clones or as a pool of many colonies.

**RNA Blot Analysis.** Total cytoplasmic RNA was isolated by Nonidet P-40-mediated cell lysis (16). The  $poly(A)^+$  RNA was fractionated by electrophoresis in 1.4% (wt/vol) agarose gels containing 2.2 M formaldehyde and transferred to nylon membranes (12). The hybridizations were carried out for 24–36 hr at 42°C in 6× NaCl/Cit (1× NaCl/Cit = 0.15 M NaCl/0.015 M sodium citrate)/0.3% nonfat dry milk/0.1% NaDodSO<sub>4</sub>/50% (vol/vol) formamide. Final washes were carried out at 55°C in 0.1× NaCl/Cit and 0.1% NaDodSO<sub>4</sub>.

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

<sup>&</sup>lt;sup>†</sup>Present address: Department of Immunology, IMM-8, Scripps Clinic, 10666 North Torrey Pines Road, La Jolla, CA 92037.

#### Immunology: Srivastava et al.



FIG. 1. PFG Southern blot hybridization of the human DNA with the RS5-specific probe. Electrophoresis of the genomic DNA from the cell line 3.1.0 was carried out in a 55-cm PFG apparatus for 72 hr at 500 V with 2-min pulses between 90° field orientations. Concatamers of  $\lambda$  vir genome (42.5 monomer) and yeast chromosomes were run as size markers. The probe 5.10 (see Fig. 2) was nick-translated to a specific activity of  $4 \times 10^8$  and hybridized in  $6 \times$  NaCl/Cit,  $1 \times$ Denhardt's solution (1× Denhardt's solution = 0.02% bovine serum albumin/0.02% Ficoll/0.02% polyvinylpyrrolidone), and 0.1% NaDodSO<sub>4</sub> for 16 hr at 68°C. Final washing of the filter was carried out with 0.1× NaCl/Cit plus 0.1% NaDodSO<sub>4</sub> three times (30 min each) at 68°C. Restriction enzymes used were *Mlu* I (lane 1), *Sfi* I (lane 2), and *Not* I (lane 3). (*Inset*) Hybridization of partial *Not* I digests of the genomic DNA with the HLA cross-reactive probe pDP001. Arrowheads indicate partial digests.

Cells were grown on Lab-tek chambers (Miles, Naperville, IL) and fixed overnight with 3% paraformaldehyde in Dulbecco's phosphate-buffered saline (PBS) at room temperature. The cells were permeabilized with 0.1% Triton X-100 (10 min) and washed with 50 mM PBS (pH 8.0) containing 20 mM lysine and 0.05% polyoxyethylene sorbitan monolaurate (wash buffer). Antiserum was applied to the cells at a dilution of 1:200 in wash buffer containing 25% goat serum and 0.3 M NaCl for 1 hr. After four changes of wash buffer, a rho-damine-conjugated F(ab')<sub>2</sub> goat anti-mouse antibody (Cooper Biomedical, Cochranville, PA) was added for 1 hr. Slides were rewashed and covered with 50% (vol/vol) glycerol/50% (vol/vol) PBS, pH 8.0.



FIG. 2. Restriction map of the cosmid clone RS5 (a) and the subclone py5.1 (b). The orientation of the gene is shown by the arrow. Regions representing the subclones 5.1A, 5.1B, and 5.10 are marked. The exon/intron organization (c) and the structure of the mature protein (d) are shown. Hatched box, class I gene-containing region; open boxes, class I cDNA-related sequence; solid box, Alu sequence. B, BamHI; H, HindIII; K, Kpn I; R, EcoRI; L, leader sequence; UT, untranslated sequence. Exons encoding the three external domains ( $\alpha$ 1,  $\alpha$ 2, and  $\alpha$ 3), transmembrane domain (TM), and cytoplasmic domains (C1, C2, and C3) are marked.



FIG. 3. RNA blot analysis of the transfected class I clones: About 40  $\mu$ g of the total cytoplasmic RNA was electrophoresed in each lane on a 1.4% agarose/2.2 M formaldehyde gel and blotted onto a nylon membrane. Hybridizations at 42°C were carried out as described using nick-translated probe pDP001 (specific activity, 10<sup>8</sup> cpm/ $\mu$ g). Lane 1 contains RNA from untransfected LMTK<sup>-</sup> cells. Transfected class I clones shown are as follows: lane 2, 10XY; lane 3, HLA-A2; lane 4, 321.2; lane 5, RS20.3; lane 6, RS5.3; lane 7, py5.1; lane 8, RS32; lane 9, RS2.1; lane 10, LN129.5.

The surface expression of the products of the transfected gees was assayed by published procedures (17). The HLA framework antibody A1.4 (9) was provided by C.-Y. Wang (Memorial Sloan-Kettering Cancer Center, New York), W6/32 was purchased from ATCC, and xenoantiserum against several B-cell and T-cell leukemias was raised in our laboratory.

### RESULTS

Identification of a Region of HLA Harboring Nonclassical Class I Type Genes. The DNA stretch telomeric to the HLA-Alocus could represent the human equivalent of the mouse Qa/TL region. Many human class I genes map to this region



FIG. 4. Cytoplasmic staining of transfected mouse cells. Cells were stained with xenoantiserum after fixing with paraformaldehyde (see text). (A) HLA-A2-transfected cells; (C) py5.1-transfected cells; (E) untransfected. (B, D, and F) Corresponding light microscopic pictures of the fluorescent stainings in A, C, and E, respectively.

1	CT88TCAT886TCTTCTTCT66AGAAACTCATGAC6CA6AGACTCAGTTCTCATTCCCA6AT66GGTCTTC66GGTTTTCTA6AGAAAGCCAATCAGC6CCAC6ACTCCCGACTCCCGAC
121	CCATCCGGACTCAAGAAGTTCTCAGGACTCAGGAGGCCGGGGTGATGGGAAGTAGGGACGGGGTGATGCGGGGCGCGCCTTCCCCAGACCTGGGGGGGG
241	BOATBBAAACGGCTTTACCGGGAGTTACACAGGGCCGGGGGGGGGG
361	TCCTTGAAGTATATTCCACTATCCTGTCCCGGCCGGGCGGG
481	ATBETECCECEGEGECCETEGATEGATEGAGEGETCAGAGEGETCAGAGTATTGEGACCEGEGAGACACEGEGAGACCECCAGEGAAACCECAAAGATTTTCCCGAGTGAATCTGCGGACCETGCGGGCCTAC
601	TACAATCABABCBABGCCG <u>BT</u> GAGTGACCCCGGCCAABGGABGCAGGTCACGACCCCCTCCCCATCCCCACGGACGGGGGGCCCCCTCGAATCTTCGGGTCCCCAGATTCACCCCAAGGC
721	Exon 3 6 S H T L Q W H H G C E L G P D G F L R G Y E Q F A Y D G K D Y L T
841	TBACTAAGGGCGGGGCCAGGGTTCCACACCCCCCCAGTGGATGCATGGCTGCGAGGCTGGGGCCCCGACGGCTTCCCCCCGCGGGTATGAACAGTTCCCCCTACGACGGCCAGGATTATCTCACC
<b>.</b>	LNEDLR <sup>®</sup> NTAVDTALQISEOKSNDASEADD0RAYLEDTCV
961	CTGAATGAGGACTGCGCCCCTGGACCGCGGCTGGACCGGCTTCAGATCTCCGAGGAAAAGTCAAATGCCTCTGAGGGGGGAGGACGACCAGGGCCTACCGGGGGGGAGAGACACATGCGTG
1081	
1201	CTCARRAGE A GARGARA CAGGGA TRAGGCCAGAA TACCACTCCTCCCTTGCATCAGGAGGGGGGGGGG
1321	CABACCTCTCTCA666666CAATTAAG6AATCTA6TCTC6CT6GA6ATTCCATCCTTCA6ATGACCT6ATGA6CAGTTCTCTTT6ACTCCCAGTATTA6GAATCAC666666AGTTTCTCTC
1441	GTGCCTGATTCTCAGCCCCCACACCAGAGTTTTTGGAGGTCTGACTCCAGCTTTTCTCAGTCACTCAGCATCCACGGCCGGC
1561	AACTCCAATAAGAAGAGGGACCCAGGACCCTAAGGTCCATTCGGTGGCAAAGGGATGGTCCTCGTCCTCTTCCTCCTACTATAATTGTCCTCTTCCTGGAGAGGAGGTGGTCACATGGGTGCT Exon 4 E P P K T H V T H H P I S D H E A T L R C W
1681	GCTGGATGTCCCATGAGAGATACAAAATGCCTTAAATTTTCTGACTCTTCCCCTCAGAGACCCCCAAAGACACCGTGACTCACCACCCCTATCTCGACCATGAGGCCACCCTGAGGTGCT
	ALGFLPAEITLTEQQDGEGHTQDTELVETRPAGDG7F0! W
1801	GOBCCCT0BGCTTCTTACCT0C0BGAGTCAACTBACCT0BCCAGCAGGGATGGGGAGGGGCCCTACCCCAGGAACACCGGAGCTCGTGGGAGCGCAGGGGATGGAAUCTTUCAGAAG
1921	BEGCAGCTGTGGTBGTGGCTTCTGGAGAGGAGGAGGAGAGAGAGAGAGAGGAGGGCATGCCGAGGCCGGAGCCCGAGCCCGAGCCCCGAGGCAGGGGAGGGGAGGGGGAGGGGAGGGGAGGGGAGGGGAGGCATGCC Exon 5 R A A S Q P T I P
2041	TCTTCTCAGGGAAAGCGGGAACCCTTCTGGAGGCCTCCGGAGGGCTGAGGGCCGGGGGGCCGGGGGCCCCTTACGTTCCCCAGAGCCGCTTCCCAGCCCACCATCC
2161	CCATCGT696CATCATT9CT69CCT66TTCTCCT166ATCT6T66TCTCT66A6CTGT66TT6CT6CT6T6ATAT66A6G6AA6A6A6CTCA66T6666A6A666GT6666TCT6
2281	AGTITICTIGICCACTGGTGTTTCAAGCCCTAGGTAAAAGTGTGCCCTGCCTCGCTACTGGGAAGACATCCACACACGAGCCTACCAGCCTGGGGCCCTGTGTCCAGCACCTACTACT
2401	TTIGTAAAGCACCTGTGACAATGAAGGACAGATTTATCACCTTGACGATTGTGGGGATGGTGGACGAGTCACAGGCAGTCACAGGGGAABGTCCCTGCTGAAGAGAGAGCCTTAGA
2521	ABBBCAST TEATCCASEACCCACACCTECTTTCTCACCTTTCTCACCTTECTEATCCTSECCTEGESTCTGCAGTBACAGTTCAGEAAACTTCTCTEGEGATCCAAAAACTAGEAGETTCCTCAGEAC
2641	
2761	CCTATAATTCCTCCTGCACCACATCTCCTGTGGGCTCTGACCAGGTCTTGTTTTTGTTCTACCCCAGGGACAGTGCCCAGGGGTCTGAGGCTCTGTAAGGCTGAAAGGCTGAAAGGCTCGAGGGCTCTGTAAAGGCTGAAAGGCTCGAAAGGCTCGAGGGACAGTGCACAGTGCCAGGGGCCCAGGGGCCCGAGGGGCCCGAGGGGCCCGAGGGGCCGCGGCG
2881	Ο 30 ΑΘΤΟ ΤΟ Τ
3001	Exon 8 (untranslated) TGACCTGAATTTGTTCATGACTATTTTCTTCTGTGCCTGAGACAGCTGCCTTGTGGCGCACTGAGATGCCAGGATTTCCTCACGCCTCCCCTATGTGTCTTAGGGGAACTCTGGCTTCTC
3121	
3241	CCTGTTCTCTTTTCTTTAAAAATAAGAACCT699GCAGAGTGCGGGCACGTCATGCCTGTAATCCCAGCACTTAGGGAGGCCCGAGGAGGGCAGATCACGAGGTCAGGAGATCGAAAACCATCC
3361	TGGCTAACACGGTGAAACCCCGTCTCTACTAAAAAAATACAAAAAATTAGCTGGGCGCACAGGCACGGGCCTGTAGTCCCAGCTACTCAGGAGGGCGGAGCAGGAGAATGGCGTCAACCCGG
3481	GAGGCGBAGGTTGCAGTGAGCCAGGATTGTGCGACTGCGCCTGCGCCGGGTGACAGGGTGAAACGCCATCTCAAAAAAATGAAAAATGAAAAAAGAACCTGGATCTCAATTTA
3601 3721	ATTTTCATATTCTTGCAATGGAATGGAACTTGAGGAAGCTTAAGGATCTAAGGACATACAGTAAATTCCACAGCACACTAGTGTAGCAAATTTAGCCTATTCTATCTCTAGCTATTCT TACACTGTTACTGACTATAACTTGAGTTGATGATGATTGTTTCTAACTGCTGTGGTGATGTCCACTACTAGTGGGATCTACTATGTGCAT 3808

FIG. 5. Complete nucleotide sequence of the class I gene contained in the clone RS5. Transcriptional promoter elements and terminal repeats of the Alu sequence are boxed. Exon/intron junctions, Alu sequence and the poly(A) adenylylation site are underlined. Amino acids are designated by the single-letter code.

(2). To identify this region at the molecular level, PFG hybridization studies were carried out using cross-reactive and locus-specific probes. The class I *HLA* cross-reactive probe pDP001 (18) identified three *Not* I fragments (1090, 540, and 190 kb) in the genome of the 3.1.0 cell line (11). The *HLA-B* locus-specific probe pRS30.4 (19) hybridized to the 1090-kb fragment, while the *HLA-A* locus-specific probe A2a.1 (20) hybridized to the 540-kb fragment. Thus, the 540-kb fragment is telomeric to the *HLA-B* locus. As evident from the partial *Not* I digests of genomic DNA (see Fig. 1 *Inset*), the 190-kb *Not* I fragment is located adjacent to the 540-kb fragment and may represent the telomeric portion of the class I region. A probe from the cosmid RS5 (5.10; Fig. 2) hybridized to this 190-kb *Not* I fragment (Fig. 1).

Expression of the Clone py5.1 on Transfection into Mouse Cell Lines LMTK<sup>-</sup> and J27. The cosmid clone RS5 cotransfected into mouse LMTK<sup>-</sup> cells expressed a very low level of detectable poly(A)<sup>+</sup> mRNA. No surface product could be detected by ELISA using the monoclonal antibodies W6/32 and A1.4 or the xenoantiserum. However, when the subclone py5.1 (Fig. 2b) was transfected, considerable amounts of class I transcript (Fig. 3) were produced, and the class I protein was detectable in the cytoplasm by immunofluoresscence (Fig. 4) using xenoantiserum. DNA from py5.1 upon transfection into the cell line, expressed high levels of a surface product detectable by ELISA either with HLA framework antibodies A1.4 and W6/32 or with the xenoantiserum (data not shown).

Complete Nucleotide Sequence and Derived Amino Acid Sequence of the Class I Gene Contained in the Clone RS5. The nucleotide sequence of the class I gene contained in the clone RS5 is shown in Fig. 5. When aligned with the nucleotide sequences of the reported class I genes (reviewed in ref. 1), an overall structural homology was observed. Two distinct features at the nucleotide level were noticeable at the 5' end. The sequence of the "TATA" box differs by a substitution of  $C \rightarrow A$  from the TCTAAA seen in previously reported human class I genes (1), and a single initiation codon is present in the gene RS5 as opposed to the two putative initiation codons in all the reported class I genes. An Alu interspersed repeat sequence (21) was found to be inserted in the middle of the 3' untranslated exon and provides a poly(A) adenylylation site. The sequence beyond this insertion does not significantly match other class I genes.

Derived amino acid sequence comparisons. The alignment of the amino acid sequences of the RS5-encoded protein and several other class I proteins is shown in Fig. 6. The average homology between RS5 and HLA-A, -B, and -C locus proteins is  $\approx 61\%$  (first external domain,  $\alpha 1$ ), 64% (second external domain,  $\alpha 2$ ), 89%  $\beta_2$ -microglobulin-binding domain,  $\alpha$ 3), 69% (transmembrane domain), and 67% (cytoplasmic domains). The average amino acid sequence homologies in the corresponding domains of the HLA-A, -B, and -C locus proteins are 80%, 81%, 93%, 70%, and 76%, respectively. The carboxyl terminus of the RS5 protein is five amino acids shorter than in most class I proteins. All of the carboxylterminal serine residues (positions 313, 314, 320, 329, 333, and 337, Fig. 6), which have been previously identified to be the sites of phosphorylation (24), are conserved. There is a single glycosylation site (Asn-Gln-Ser) in the  $\alpha$ 1 domain at residues 85-87. Cysteine residues at positions 101 and 164 (in the  $\alpha 2$  domain) and positions 203 and 259 (in the  $\alpha 3$  domain) are conserved.

FIRST DOMAIN

RS5 HLA-A2 HLA-A3 HLA-AW24 HLA-A28 HLA-B7 HLA-B27 HLA-CW3 HLA-CW3 HLA-12.4	l GSHSI I C R	LKYIPLSCP MR FFT VS MR FFT VS MR FYT VS MR FYT VS MR FHT VS MR FCTAVS MR FCTAVS MR FYTTMS	GPA-GEPRF R GR R GR R GR R GR R GR R GR R GR R	TISVGYVC A A A A T A	ETQFVR D D R D D D D D D D D D D	FDNDAAS S S S S S S DE S D S D	PRMVPRAF Q E Q E E E E E E E E E E E E E E E E E	WMEQEGSEYW I P I P I P I P V RK P V RK K	DRETRSARET G KVKAH Q NVKAQ E GKVKAH N NVKAQ N QIYKAQ N QIYKAQ QKYKPQ QKYKRQ N QICKAQ	AKIFRVNLR SQTH D G SQTD D G SQTD D G QTD ES QTD ED QTD S QAD S QTE E	9C RTVRAYYNQSDT G L G EA I L G EA I L G EA I L G EA NL G EA NL G EA NL G D I ALR G	
SECOND DOMAIN												
RS5 HLA-A2 HLA-A3 HLA-A24 HLA-A28 HLA-B7 HLA-B27 HLA-B27 HLA-CW3 HLA-328 HLA-12.4	91 GSHTLQU VF II L A S A N II R YR M V	IMHGCELGP Y DV S Y DV S Y DV S Y DV S Y DV S Y DV Y DV S D Y DV Y DV Y DV	DG-FLRGYE WR H R RL HU RL D RL D RL D RL D RL D	QFAYDGK Y D Y D Y D H S H	DYLTLNE IA K IA K IA K IA K IA IA IA IA	EDLRSWT. S	AVDTALQI AMA AMA AMA A-MA AMA AMA AAA AAA AAAA	SEQKSNDASEI TKH WEA HV TRK WEA HV TKR WEA HV TKH WEA HV TQR WEA RV TQR WEA RV TQR WEA RV TQR LEA RA TKR WEA RR	ADDQRAYLED EQL G EQL DG EQ G EQ G EQR G EQL G EQL G EQL K EQK K G EQR V G	TCVEWLHKY RR DG RR -R- E RR E RR KRR I EF RR	182 LEKGKETLLHL N QRT N QRA N QRA N DK ERA N QRV KN QQG K N KQRA N QRA	
THIRD DOM RS5 HLA-A2 HLA-A3 HLA-A28 HLA-A28 HLA-B27 HLA-203 HLA-228 HLA-328 HLA-12.4	AIN 183 EPPKTHV DA D M D M D M D M D M	THHPISDHE AV - AV V L	H	GFLPAEI 5 Y Y Y Y Y Y Y Y	FLTEQQD WR WR WR WR WR WR WR WR WR WR	GEGHTQI DQ DQ DQ DQ DQ DQ DQ DQ DQ DQ DQ DQ	DTELVETRI	PAGDGTFQKWA V R R	AVVVPSGEEC Q Q Q	)RYTCHVQHE M	274 EGLPEPVTLRW K L K L K L K L K L C L Q L S L S	
TRANSMEMBP RS5 HLA-A2 HLA-A3 HLA-B3 HLA-B27 HLA-B27 HLA-288 HLA-12,4	ANE AND CY 75 EAASQPTIP EPS ELS EPS EPS EPS EPS EPS EPS EPS EPS V	TOPLASMIC IVGIIAGLV V PA V PA V A V A V A V A	DOMAINS FAIT AIT VAVI VAVI VAVI VAVI VAVL-L V VVLAL VAT	AVVAAVI M M M M M M V M T M	WRKKSSO R 1 R 1 R R 1 C R C R C R C R C R C R C R C R	GGKGGSY )R JR JR JR C JR C JR Y	SKAEGSDS Q AS TQ AS Q AC Q AC Q AC Q AS N Q AC N Q A N	AQGSESHSL- DV- T DV- T DV- T DV- T DV- T DV- T DV- T DE- I DV- T	343 ACKV ACKV ACKV A ACKA ACKA A			

FIG. 6. Comparison of nucleotide sequence-derived amino acid sequence of clone RS5 with other class I proteins. The sequences of HLA-A2, A3, AW24, A28, CW3, 328, and 12.4 are as compiled by Klein and Figueroa (22), and the sequence of HLA-B27 is from ref. 23. Amino acids are designated by the single-letter code.

Davidson *et al.* (25) have identified five variable regions in the class I proteins. These regions span amino acids 62–80 ( $\alpha$ 1 domain), 95–116 and 137–163 ( $\alpha$ 2 domain), 177–194 (shared between  $\alpha$ 2 and  $\alpha$ 3), and 293–300 in the transmembrane domain. The majority of the amino acid substitutions are confined to these variable regions in the RS5 protein (Fig. 6). RS5 has an additional divergent region in the  $\alpha$ 1 domain spanning residues 5–16, corresponding to the first variable segment of class I proteins identified by Sood *et al.* (26). The hydrophobic character of the transmembrane domain and the hydrophilic nature of the cytoplasmic domain are maintained.

*Expression of RS5 in JY cell line*. As shown in Fig. 7, a 1600-nucleotide-long mRNA transcript could be detected in the B-cell line JY when the 30-nucleotide-long probe (complementary to nucleotide residues 3239–3267 in Fig. 5) specific for the gene *RS5* was used. Little or no transcript was detected in the T-cell lines MOLT-4, HuT 78, HuT 102, and SU. No transcript was detected in the B-cell line 3.1.0, from which the clone RS5 was derived. No detectable hybridization was seen with the cloned genes from the *HLA-B* and *-A* loci (Fig. 7).

#### DISCUSSION

The gene sequences of several alleles of the polymorphic loci HLA-A, -B, and -C have been reported (reviewed in ref. 1). While the structure, mapping, and function of the class I genes of these three loci are well documented, little is known about the structure, mapping, and function of an additional 20 or more genes belonging to the class I family, except that

some are pseudogenes. Many, but not all (17), of these additional class I genes map telomeric to the *HLA-A* locus.

Of a large number of cloned class I genes other than HLA-A, -B, or -C transfected into mouse LMTK<sup>-</sup> cells, five made full-length transcripts detectable with the human class I cross-reactive probe pDP001 and also produced serologically detectable peptides. One of these clones, RS5, was subjected to detailed analysis for the following reasons: (i) it was one of a small group of genes contained within a large EcoRI fragment (most class I genes reside in 6- to 7-kb EcoRI fragments), (ii) the gene RS5 was present in each of two genomes examined, (iii) the gene RS5 was located in the telomeric region of the major histocompatibility complex.

The gene has a similar pattern of exon/intron organization to known class I genes. No remarkable differences in the lengths of the introns are noticed such as have been found between H2 and Qa/TL genes (27). As with the H2 D/L (28) and TL (29) region genes, but unlike HLA-A, -B, or -C genes, a full-length SINE (21) element was found at the 3' untranslated end of the gene RS5. This is cotranscribed and provides the poly(A) adenylylation site.

Conserved amino acid residues, resembling those in many of the mouse or rabbit class I proteins (22), can be found in the *RS5*-encoded protein where it diverges from the human class I proteins. Such amino acid residues are seen at position 10 (Leu), 16 (Ala), 66 (Ala), 69 (Thr), 72 (Ile), and 73 (Phe) in the first domain; residues 97 (Trp), 102 (Glu), 103 (Leu), 114 (Glu), 116 (Phe), 136 (Val), 147 (Ser), 169 (His), and 180 (Leu) in the  $\alpha 2$  domain; and residues 334 (Glu) and 326 (Gly) in the cytoplasmic domains (see Fig. 6). The level of amino acid



FIG. 7. RNA blot analysis of human cell lines using 30-nucleotide-long oligomer probe specific for RS5. About 5  $\mu$ g of poly(A)<sup>+</sup> RNA was electrophoresed in lanes 1–3 and 40  $\mu$ g of total cytoplasmic RNA was used in lanes 4 and 5. The oligomeric probe was labeled to a specific activity of 3 × 10<sup>8</sup> cpm/ $\mu$ g. Hybridization was carried out at 42°C in 6× NaCl/Cit. The blot was washed with 2× NaCl/Cit and 0.1% NaDodSO<sub>4</sub> at 55°C. Lanes: 1, JY; 2, MOLT-4; 3, SU; 4, HuT 78; 5, HuT 102. (*Inset*) Dot blot hybridization of the same oligomeric probe with the DNA from clones py5.1 (A), JY150 (HLA-B7; B),  $\lambda$ 307 (HLA-A2; C), and RNAs from the cell lines 3.1.0 (D), L/RS5 (py5.1-transfected mouse cells; E and F), JY (G), MOLT-4 (H), HuT 78 (I), and HuT 102 (J).

sequence divergence between the three external domains of RS5 and *HLA-A*, -B, and -C locus proteins is of the same order as noticed between H2 and several of the Qa/TL region gene products (6).

In the mouse, the Qa/TL region has been molecularly characterized and represents an array of relatively nonpolymorphic class I genes (3). Certain of these genes encode B-cell-, T-cell-, and leukemia-specific membrane molecules. Others are expressed as secretory proteins (30). No function has been assigned to these murine gene products and they do not appear to serve as restriction elements. Some of the genes in this region have been identified as a potential source of donor sequences involved in the gene conversion events (31). Possible human analogues of mouse Qa- and TL-like proteins (32) have been identified based on the selective expression in populations of lymphoid cells of certain  $\beta_2$ -microglobulinassociated proteins (33). The genes encoding putative human TL equivalent antigens (CD1) have been cloned (34) and found to constitute a family of five genes that has been mapped outside of the HLA complex (on chromosome 1). The restricted expression and the molecular mapping of RS5 are consistent with the possibility that it may be the human analogue of a murine Qa/TL region gene. It remains to be determined whether there is any functional distinction between RS5<sup>+</sup> and RS5<sup>-</sup> cells.

The authors wish to thank Dr. V. B. Reddy and Dr. H. Orr for providing the vector poly-ori-Bam and the HLA-A2-specific probes, respectively. We appreciate the excellent technical assistance of Yurico Yamakawa, Stephanie White, and Susan Pleasic. We are grateful to Ann M. Mulvey for preparing this manuscript. This work was supported by grants from the National Cancer Institute.

- Srivastava, R., Duceman, B. W., Biro, P. A., Sood, A. K. & Weissman, S. M. (1985) Immunol. Rev. 85, 93-121.
- 2. Orr, H. T. & DeMars, R. (1983) Nature (London) 302, 534-536.
- Hood, L., Steinmetz, M. & Malissen, B. (1983) Annu. Rev. Immunol. 1, 529-568.
- Malissen, M., Malissen, B. & Jordan, B. R. (1982) Proc. Natl. Acad. Sci. USA 79, 893–897.
- Biro, P. A., Pan, J., Sood, A. K., Kole, R., Reddy, V. B. & Weissman, S. M. (1983) Cold Spring Harbor Symp. Quant. Biol. 42, 1082-1086.
- Fisher, D. A., Hunt, S. W., III & Hood, L. (1985) J. Exp. Med. 162, 528-545.
- Gladstone, P. L., Fueresz, L. & Pious, D. (1982) Proc. Natl. Acad. Sci. USA 79, 1235–1239.
- Terhorst, C., Parham, P., Mann, D. L. & Strominger, J. L. (1976) Proc. Natl. Acad. Sci. USA 73, 910–914.
- Bushkin, Y., Posnett, D. N., Pernis, B. & Wang, C.-Y. (1986) J. Exp. Med. 164, 458-473.
- Kavathas, P. & Herzenberg, L. A. (1983) Proc. Natl. Acad. Sci. USA 80, 524-528.
- Lawrance, S. K., Smith, C. L., Srivastava, R., Cantor, C. & Weissman, S. M. (1986) Science 235, 1387–1390.
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
- Lawrance, S. K., Das, H. K., Pan, J. & Weissman, S. M. (1985) Nucleic Acids Res. 13, 7515-7528.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
- Wigler, M., Sweet, R., Sim, G. K., Wold, B., Pellicer, A., Lacy, E., Maniatis, T., Silverstein, S. & Axel, R. (1979) Cell 16, 777-785.
- 16. Favaloro, J., Triesman, R. & Kamen, R. (1980) Methods Enzymol. 65, 718-720.
- Duceman, B. W., Ness, D., Rende, R., Chorney, M. J., Srivastava, R., Greenspan, D. S., Pan, J., Weissman, S. M. & Grumet, F. C. (1985) Immunogenetics 23, 90-99.
- Sood, A. K., Periera, D. & Weissman, S. M. (1981) Proc. Natl. Acad. Sci. USA 78, 616–620.
- Srivastava, R., Lawrance, S. K., Smith, C. L., Cantor, C. R. & Weissman, S. M. (1986) Trans. Assoc. Am. Physicians 99, 1-12.
- 20. Keller, B. H. & Orr, H. T. (1985) J. Immunol. 134, 2727-2733.
- Duncan, C. H., Jagadeeswaran, P., Wang, R. C. & Weissman, S. M. (1981) Gene 13, 185-196.
- 22. Klein, J. & Figueroa, F. (1986) Immunology Today 7, 41-44.
- 23. Szots, H., Riethmullar, G., Weiss, E. & Meo, T. (1986) Proc. Natl. Acad. Sci. USA 83, 1428-1432.
- Guild, B. C. & Strominger, J. L. (1984) J. Biol. Chem. 259, 13504–13510.
- Davidson, W. F., Kress, M., Khoury, G. & Jay, G. (1985) J. Biol. Chem. 260, 13414-13423.
- Sood, A. K., Pan, J., Biro, P. A., Pereira, D., Srivastava, R., Reddy, V. B., Duceman, B. W. & Weissman, S. M. (1985) *Immunogenetics* 22, 101-121.
- Ronne, H., Widemark, E., Rask, L. & Peterson, P. A. (1985) Proc. Natl. Acad. Sci. USA 82, 5860-5864.
- Kress, M., Barra, Y., Khoury, G. & Jay, G. (1984) Science 226, 974–977.
- Pontarotti, P. A., Mashimo, H., Zeff, R. A., Fischer, D. A., Hood, L., Mellor, A., Flavell, R. A. & Nathenson, S. G. (1986) Proc. Natl. Acad. Sci. USA 83, 1782–1786.
- Devlin, J. J., Lew, A. M., Flavell, R. A. & Coligan, J. E. (1985) EMBO J. 4, 369-374.
- Nathenson, S. G., Galiebter, J., Pfaffenbach, G. M. & Zeff, R. A. (1986) Annu. Rev. Immunol. 4, 471-502.
- 32. Flaherty, L. (1981) in Role of the Major Histocompatibility Complex in Immunobiology, ed. Dorf, M. (Garland, New York), pp. 35-57.
- Kahn-Perles, B., Wietzerbin, J., Caillol, D. H. & Lemonnier, F. A. (1985) J. Immunol. 134, 1759–1765.
- Martin, L. H., Calabi, F. & Milstein, C. (1986) Proc. Natl. Acad. Sci. USA 83, 9154-9158.