Immunogenetics (1998) 48: 22–31
© Springer-Verlag 1998

ORIGINAL PAPER

Shigeo Katabami · Akihiro Matsuura · Hong-Zhi Chen
Kohzoh Imai · Kokichi Kikuchi

Structural organization of rat *CD1* typifies evolutionarily conserved *CD1D* class genes

Received: 29 October 1997 / Revised: 9 December 1997

Abstract The non-major histocompatibility complex (MHC)-encoded CD1 family has recently emerged as a new antigen-presenting system that is distinct from either MHC class I or class II molecules. In the present study, we determined the genomic structure of the rat *CD1* locus. It was extremely similar to mouse *CD1* genes, especially to *CD1D1*. The 5′ flanking region of the *CD1* gene contained the binding motifs for two cytokine-inducible transcription factors, NF-IL2-A and NF-IL6. Some regulatory elements found in MHC class I genes (enhancer A, enhancer B, and the IFN response element) were absent. It is of interest that a tyrosine-based motif for endosomal localization found in the human CD1b cytoplasmic tail was encoded by a single short exon which was conserved in all CD1 molecules except for CD1a. Southern blot and direct sequencing analyses of inbred rat strains suggested very limited polymorphism in the 5′ region where a hydrophobic ligand-binding groove is encoded; a single base substitution resulted in amino acid alteration of alanine (GCT) to valine (GTT) at codon 119. Comparison of the overall exon-intron organization of *CD1* genes revealed that the length of the intron was also characteristic to each of the two classes of *CD1* genes, classic *CD1* and *CD1D*; such categorization has hitherto been made according to the sequence similarity of the coding region. This finding provides further support for the hypothesis that the two classes have different evolutionary histories. In contrast to the complete absence of the classic *CD1* in rats and mice, the entire region of nonpolymorphic *CD1D* has been conserved through mammalian evolution. Similar functional properties of rodent CD1 and human CD1d are implied.

Key words CD1 · Rat · Gene · Organization · Polymorphism

The nucleotide sequence data reported in this paper have been submitted to the EMBL/GenBank/DDBJ nucleotide databases and have been assigned the accession number AB002172

S. Katabami · A. Matsuura (✉) · H.-Z. Chen · K. Kikuchi
Department of Pathology, S1 W17, Sapporo Medical University, School of Medicine, Chuo-ku, Sapporo 060, Japan

K. Imai
Department of Internal Medicine, S1 W17, Sapporo Medical University, School of Medicine, Chuo-ku, Sapporo 060, Japan

Introduction

CD1s were the first human leukocyte differentiation antigens defined by monoclonal antibodies with structural similarity to major histocompatibility complex (MHC) class I molecules (McMichael et al. 1979). They are expressed on the cell surface as a heterodimer of an α-chain of approximately 45 000 M_r non-covalently associated with β_2-microglobulin (β_2m) (Boumsell 1989; Calabi et al. 1991). In contrast to MHC class I, they show limited polymorphism and do not map to the MHC. In humans, there are five distinct *CD1* loci: *CD1A, B, C, D,* and *E* (Calabi and Milstein 1986; Calabi et al. 1989b; Martin et al. 1986). These encode the serologically defined CD1a, b, c, and d molecules. Sequence comparison of the leader, α1, and α2 domains has allowed categorization of the *CD1* genes into two distinct classes: the "classic CD1 class" (*CD1A, -B, -C*) and the "CD1D class" (*CD1D*) (Calabi et al. 1989b; Hughes and Nei 1991). The protein products of these gene classes, also referred to as group I and group II proteins, respectively, have different tissue distributions (Porcelli 1995). Three classic CD1 molecules are expressed not only on immature cortical thymocytes but also on professional antigen-presenting cells such as epidermal Langerhans cells, dermal dendritic cells, and cytokine-activated monocytes (Cattoretti et al. 1989; Kasinrerk et al. 1993; Porcelli et al. 1992; Teunissian et al. 1990). By contrast, the CD1d molecules are abundantly expressed by nonlymphoid organs such as liver and lymphoid organs, whereas the expression of mouse CD1 by intestinal epithelial cells is controversial (Blumberg et al. 1991; Brossay et al. 1997; Canchis et al. 1993).

The *CD1* family genes have also been found in many mammalian species such as mice, rats, rabbits, and sheep (Bradbury et al. 1988; Calabi et al. 1989a; Ferguson et al. 1996; Ichimiya et al. 1994). In most cases, exonic sequences were characterized by cDNA cloning. On the other hand, the complete genomic structures including intronic sequences are only available for a few *CD1* genes, such as *CD1D* and *CD1E* (Calabi et al. 1989b). We have previously shown that rat CD1 belongs to the second CD1D class and is expressed by a wide variety of cells (Ichimiya et al. 1994; Kasai et al. 1997; Matsuura et al. 1997). In the present study, we isolated genomic clones encompassing the rat CD1 locus and determined the sequences to characterize rat *CD1* in detail and gain an insight into the CD1D class. Comparison of the structural organization of the rat *CD1* gene with those of previously characterized *CD1* family genes was conducted. The extent of rat CD1 polymorphism was also investigated by restriction fragment length polymorphism (RFLP) and direct sequencing analyses using inbred rat strains. The functional significance of evolutionarily conserved CD1d protein is discussed.

Materials and methods

Screening of a rat genomic library

An F344 rat liver genomic library constructed in the lambda DASH vector was obtained from Strategene (La Jolla, Calif.). Transfection into XL1-Blue MRA (P2) competent cells resulted in a total of 10^6 independent plaques. Approximately 4×10^5 plaques were plated onto 15-cm dishes and transferred to colony/plaque screen filters (NEN, Boston, Mass.). Before hybridization, the filters (NEN) were incubated at 42 °C for over 16 h in a buffer containing 0.25% BSA, 0.25% polyvinyl-pyrrolidone, 0.25% Ficoll (type 400; Pharmacia Japan Tokyo, Japan), 62.5 mM Tris-HCl (pH 7.5), 0.125% sodium pyrophosphate, 1.25% sodium dodecyl sulfate (SDS), 12.5% dextran sulfate, 50% formamide, and 100 μg/ml denatured salmon sperm. Then the filters were hybridized with a full-length rat CD1 cDNA probe (27.1, Ichimiya et al. 1994) radiolabeled by the random hexamer priming method at 42 °C for 24 h in the same buffer. Washing was started at 65 °C for 20 min in 2 × standard sodium citrate (SSC), 0.1% SDS, and then the salt concentration was decreased stepwise to 0.2 × SSC. Positive plaques were re-cloned.

Restriction mapping, DNA sequencing, and primer extension

Restriction enzymes were obtained from New England Biolab (Boston, Mass.). Cloned DNA was digested and subcloned into pBluescript II SK$^-$ vector. Double-stranded plasmid DNA was sequenced with the ABI373A DNA Sequencer using dye terminator cycle sequencing with Ampli*Taq* DNA Polymerase, FS, according to the protocols included with the kit (ABI, Foster City, Calif.). The sequence run was carried out on 6% polyacrylamide gels.

Primer extension was performed according to the standard procedure (Sambrook et al. 1989), with minor modifications. Briefly, a primer inversely complementary to exon 2 (e.g., N292:5′-AA-CAGGGGTCTTGACACCCTTACGGGTGTC-3′), was 5′ end-labeled with [γ-^{32}P] ATP (3000 Ci/mmol) and T4 polynucleotide kinase. Ten picomoles of ^{32}P-labeled primer was co-precipitated with 50 μg of total RNA or 5 μg of mRNA isolated from the thymus and the liver of five-week-old F344/Crj rats (Fischer, RT1 haplotype; *lv1*). After hybridization at 37 °C, the primer was extended with Moloney mouse leukemia virus reverse transcriptase (Seikagaku, Tokyo, Japan) under the conditions recommended by the manufacturer. After phenol/chloroform extraction and ethanol precipitation, the pellet was resuspended in a formamide dye buffer and denatured by boiling for 5 min. The samples were analyzed on an 8% denaturing sequencing gel.

Southern blot analysis

The restriction enzyme digests of cellular DNA obtained from eleven different rat strains (Aizawa and Natori 1988); F344/Crj (RT1 haplotype, *lv1*), LEW/Hkm (*l*), Wistar/Smc (*l*), NIGIII (*q*), LEJ/Hkm (*j*), ALB/Hok (*b*), BN/Hok (*n*), ACI/Hkm (*av1*), TO/Hkm (*u*), WKAH/Hkm (*k*), and W/N/Hkm (*k*), were separated on 0.7% agarose gel and transferred to a Gene Screen Plus blotting membrane (NEN) by the capillary transfer method. The blots were prehybridized and then hybridized with a 5′ probe generated by polymerase chain reaction (PCR) amplification of the F13.2S genomic clone using two rat CD1-specific oligonucleotides, N366 (5′-TCGGAGCCCAGGGCTGTG-TAGA-3′) from the 5′ untranslated (UN) region and N392 (5′-GCAGGTGTCGTTCAGGAG-3′) complementary to near the end of exon 3. The probe was labeled with α^{32}P-dCTP, added to a minimum volume of hybridization solution, and incubated with the blot for 18 h at 42 °C. Washing was decreased to 1 × SSC.

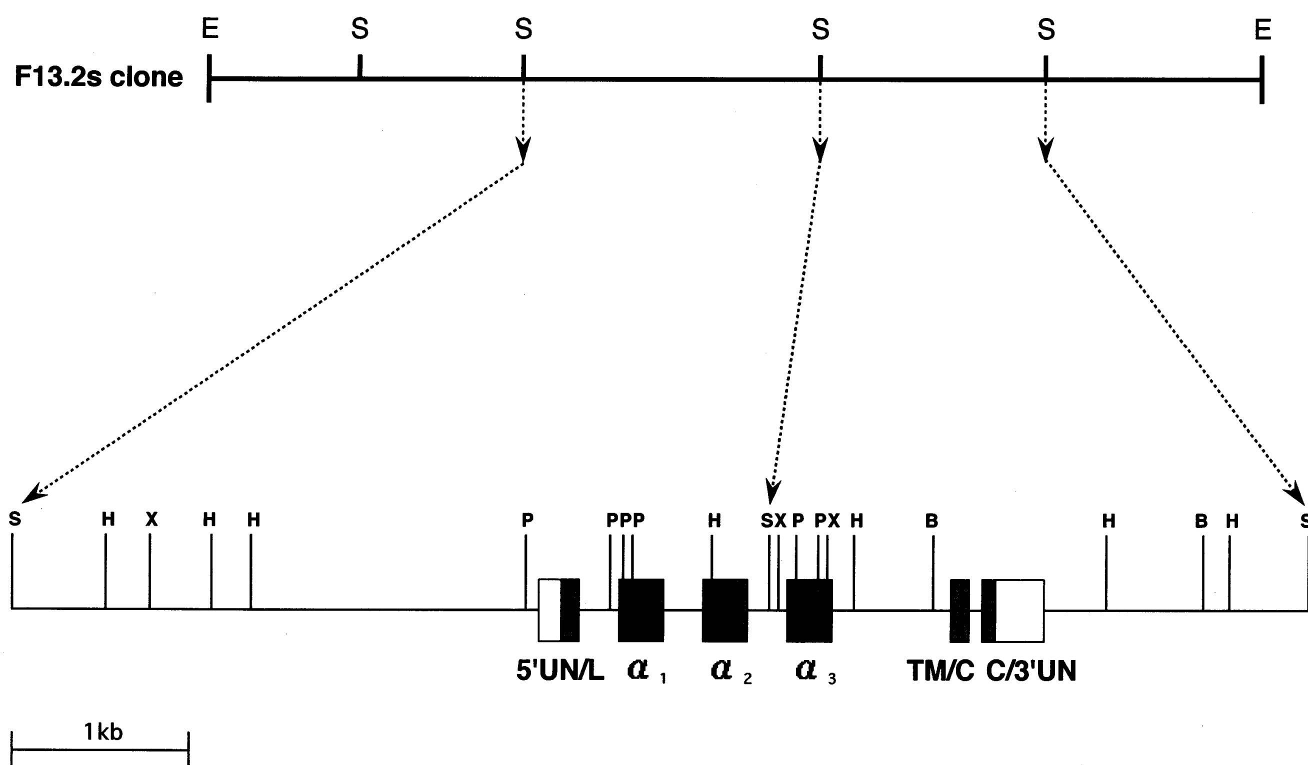
Polymerase chain reaction and direct sequencing analysis

Cellular DNAs from SDJ/Hok (*u*) and the eleven rat strains mentioned above were subjected to PCR as described previously (Itoh et al. 1993) using two rat CD1-specific oligonucleotides designated M1 and M2. M1 (5′-CCTGCAGTCTATCTGCTG-3′) corresponds to the area near the 3′ end of intron 1, and M2 (5′-AGATGGATCCAAGTGGAGAA-3′) is inversely complementary to the 5′ end sequence of intron 3. The PCR conditions were 30 cycles of 30 s at 94 °C, 30 s at 55 °C, 1 min at 72 °C, and a final extension of 7 min at 72 °C. Following this step, nested PCR was carried out using one microliter of amplified products wtih a pair of primers; M4 (5′-ATCTGCTGATTCGCTATG-3′) corresponding to the 3′ end sequence of intron 1 and M3 (5′-GTCGAGGTTCTGGATCACTC-3′) inversely complementary to the 5′ end sequence of intron 2. This yielded a DNA fragment containing exon 2 coding for the α1 domain. Nested PCR was also done with another pair of primers; M5 (5′-CTATAATCTTCATGCCAACT-3′) corresponding to the 3′ end sequence of intron 2 and M2. The latter reaction yielded a DNA fragment containing exon 3 coding for the α2 domain. The conditions of PCR were the same as mentioned above. Each PCR product was purified using Microcon-100 microconcentrator (Amicon, Beverly, Mass.). Both sense and anti-sense strands were sequenced using nested PCR primers.

Results

Isolation and characterization of rat CD1 gene

To isolate the genomic clones encoding rat CD1, we screened the F344 rat genomic library using a full-length rat CD1 cDNA (27.1) as a probe. Of three overlapping genomic clones isolated from a total of 5×10^6 plaques, a clone designated F13.2S with a longest insert approximately 17 kilobases (kb) in length was characterized in detail. Restriction mapping and DNA sequence analysis showed that the rat *CD1* gene spanned an approximately 4 kb region of DNA (Fig. 1). No other *CD1* gene was found within 7 kb upstream and downstream of the coding region (Fig. 1 and data not shown), confirming a previous assumption based on Southern blot analysis that rat *CD1* is a single-copy gene.



Precise exon-intron boundaries of the rat *CD1* gene were established by comparing cDNA and genomic sequences (Fig. 2). The genomic sequences differed from the 27.1 cDNA sequences in four nucleotide positions, two of which were located in exon 2, while the others were in the 3′ UN region. Since cDNA and genomic clones were isolated from the same F344 rat strain and no such polymorphism was found in the corresponding positions of different rat strains as described in the following section, we re-sequenced the entire 27.1 cDNA clone by the automated cycle sequencing method which was newly adapted for this genomic sequencing analysis. We confirmed that the genomic sequences were correct; in exon 2, T at nucleotide position 179 in 27.1 (C<u>T</u>G: Cys) was actually G (C<u>G</u>G: Arg) and C at 365 (<u>C</u>AT, His) was T (<u>T</u>AT, Tyr); in the 3′ UN region, C at 1167 was T and T at 1376 was A.

The rat *CD1* gene contained six exons encoding the following regions or domains (Figs. 1, 2); exon 1, the 5′ UN region and the leader peptide; exons 2–4, the α1–3 domains, respectively; exon 5, the transmembrane region and part of the cytoplasmic tail; and exon 6, the remainder of the short cytoplasmic tail and the 3′ UN. Therefore, the overall exon-intron organization of the rat *CD1* gene was similar to that of the MHC class I gene, in that individual functional domains were encoded by separate exons. The nucleotide sequences surrounding the exon-intron boundaries all conformed to the GT/AG rule. As is usually the case with MHC class I and class II genes, RNA splicing always took place between the first and second bases of the junctional codons with the exception of those of exon 5 and exon 6 (Figs. 2, 3).

Fig. 1 Organization and restriction map of the rat *CD1* gene. Six exons shown as *boxes* are based on 27.1 cDNA sequences. Amino acid coding regions are *shaded* and 5′ and 3′ UN regions are *not shaded*. Restriction sites are indicated as *E*, *Eco* RI; *S*, *Sac* I; *H*, *Hin* dIII; *X*, *Xba* I; *P*, *Pst* I; *B*, *Bam* HI. *5′*UN, 5′ untranslated region; *L*, leader; α1–α3, exons encoding α1–α3 domains; *TM*, transmembrane region; *C*, cytoplasmic region; *3′*UN, 3′ untranslated region

The cytoplasmic portion of the CD1 molecule was encoded by the 3′ end of the fifth exon and short sixth exon; a stretch of charged amino acid residues, RRR in rat CD1, was followed by consensus sequence YQXI/V, YQDI in rat CD1. These features were well conserved in all CD1 molecules except for CD1a (Fig. 3). The tyrosine-based motif, YQNI, was recently reported to be a signal for internalization and targeting of human CD1b antigens to endosomal compartments (MIICs, MHC class II compartments) where antigen-loading of class II molecules by endocytosed peptide occurs (Sugita et al. 1996). CD1a molecules lost this motif by a point mutation at the codon for Q (CAA) to a stop codon (TAA) in the sixth exon, indicating that prototypical CD1 might carry this motif and be used for trafficking to the endocytic system.

Structural features of the 5′ flanking region of the rat CD1 gene

For reasons that are not understood, trials for determination of the transcription initiation site of rat *CD1* by primer extension analysis using several primers complementary to the exon 2 sequence, N292 (5′-AACAGGGGTCTTGA-CACCCTTACGGGTGTC-3′) and other primers (data not

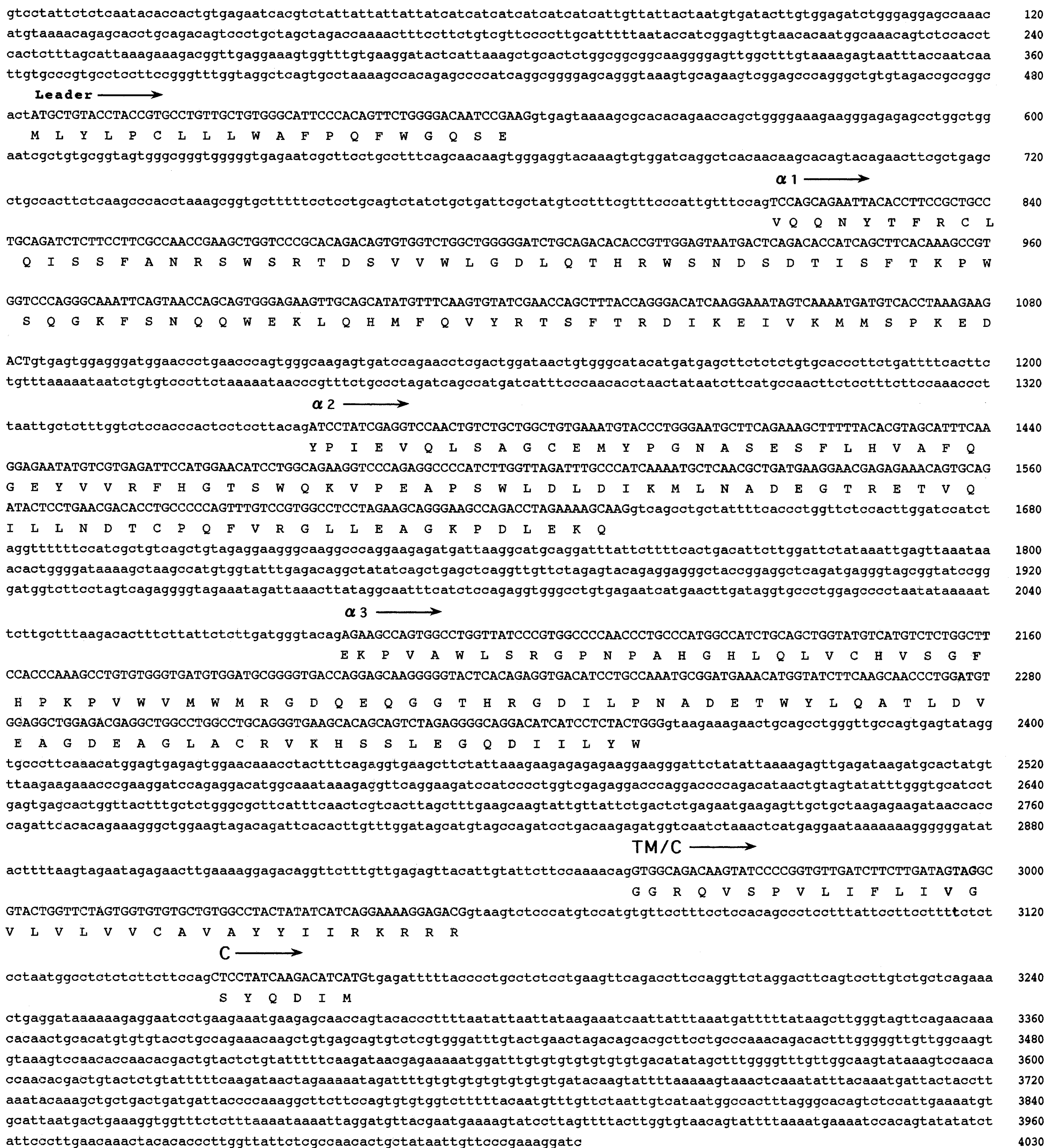


Fig. 2 Nucleotide sequence of the rat *CD1* gene. Exonic and intronic sequences are written in *capital* and *lowercase* letters, respectively. Nucleotide position 1 was the first sequence determined in this study. Deduced amino acid sequences are shown *below* the nucleotide sequences. The leader peptide, extracellular (α1, α2, and α3) domains, transmembrane (TM) and cytoplasmic (C) regions are indicated by *arrows*. The CCAAT box is at 353–357. A polyadenylation signal is at 3873–3878

shown) failed, as previously reported for human CD1s (Calabi et al. 1991). We therefore used the 5′ UN region of 27.1 cDNA as a guide to examine the sequences from base 1 to base 386 shown in Fig. 2, as 27.1 starts from nucleotide position 387. A CCAAT box, one of the obvious promoter sequence elements, was found at position 353 [131 base pairs (bp) upstream of the translational start site]. However, TATA and CCGCCC boxes were not found. A conserved nonamer (AATCTTGGA) and two heptamers

```
             exon 5                     intron 5                          exon 6
        <------------------|---------------------------|-------------------------------------->

              R   R   Ar                                    g  S  | Y   Q   D   I  | M   *
rat CD1      AGG AGA CG- -- gtaagtctcc   69bp    cttcttccag C TCC |TAT CAA GAC ATC |ATG TGA

              R   R   R   Se                                r  A  | Y   Q   D   I  | R   *
mCD1d1       AGA AGG AGA AG gtaagtct     ND        tcttccag C GCT |TAT CAA GAC ATC |CGG TGA

              R   K   R   Cy                                s  F  | C   *          |
hCD1a        AGG AAA CGC TG gtgagttctt  (169bp NA) tctcatccag T TTC |TGT TAA GAC ACA |CCA TGA

              M   R   R   Ar                                g  S  | Y   Q   N   I  | P   *
hCD1b        ATG AGG CGC CG gtgagttggt  (553bp NA) tttttaacag G TCA |TAT CAG AAT ATC |CCA TGA

              K   K   H   Cy                                s  S  | Y   Q   D   I  | L   *
hCD1c        AAG AAG CAC TG gtga         ND                 C TCA |TAT CAG GAC ATC |CTG TGA

              K   R   Q   Th                                r  S  | Y   Q   G   V  | L   *
hCD1d        AAG AGG CAA AC gtaagttctc   75bp    tctctcacag T TCC |TAT CAG GGC GTC |CTG TGA

                                  endosomal localization motif      Y   Q   X  I/V
```

Fig. 3 Comparison of nucleotide and amino acid sequences of the cytoplasmic portion of CD1. *mCD1D2* is the same as *mCD1D1*. A conserved consensus sequence, YQXI/V, shown by a *box*, is encoded by the sixth exon. Absence of this motif in CD1a is due to a non-sense mutation in the first nucleotide for codon Q. *ND*; not determined. *NA*; not available from the GenBank/EMBL/DDBJ database

Table 1 Potential transcriptional regulatory motifs in the 5′ flanking region of the rat *CD1* gene

Transcription factor	Consensus sequence[a]	Sequence in the gene[b]	Nucleotide position[c]	Mismatch[d]
E2A	RCAGNTG	CACCTGC (rev)	253	0
GATA-1	MYWATCWY	ATTATCAT	338	0
NF-IL2-A	ATGTAAAACA	ATGTAAAACA	266	0
NF-IL6	TKNNGNAAK	TGAGGAAAG	119	0
TCF-1	MAMAG	CTGTG (rev)	364	0
		AACAG	260	0
		AACAG	160	0
		CTTTG (rev)	54	0
		AAAAG	48	0

[a] The single-letter codes for ambiguous bases are as follows: R = A or G; Y = C or T; M = A or C; K = G or T; N = any; W = A or T
[b] "rev" indicates that the DNA motif is on the opposite strand
[c] Since the transcription initiation site could not be defined, the number is based on the sequences in Fig. 2 determined in this study and refers to the first nucleotide in question
[d] Number of bases that do not correspond to the consensus sequence

(GGGAAAT and GAAGTCA) were previously identified upstream of the translational start sites in human *CD1A, B, C,* and *E* (Calabi et al. 1989b). Like human *CD1D*, rat *CD1* did not contain such sequence elements. A computer-assisted search of D. Ghosh's transcription factor database (Release 7.0) allowed the identification of several consensus *cis*-acting regulatory elements (Table 1). Of note was the existence of binding motifs for NF-IL2-A and NF-IL6, which are known to be induced by IL-2 and IL-6, respectively (Kamps et al. 1990; Majello et al. 1990). The 5′ flanking region of the MHC class I gene contains regulatory sequences known as enhancer A, enhancer B, and the IFN response element (Ting and Baldwin 1993). None of these sequences was found in the 5′ flanking region of the rat *CD1* examined in the present study.

Comparative analysis of structural organization of CD1 family genes

We compared rat CD1 exonic and intronic sequences with those of other *CD1* genes. The percent similarity and length of each exon and intron are shown in Table 2. Although a number of *CD1* genes have been identified, sequences available from public databases are mainly for exons, since most studies were performed by cDNA cloning. Therefore, we collected useful information about introns and untranslated regions from previously published papers (Aruffo and Seed 1989; Balk et al. 1989; Balk et al. 1991; Blumberg et al. 1991; Bradbury et al. 1988; Calabi et al. 1989; Martin et al. 1986). While the protein coding regions of rat *CD1* were equally similar to two mouse *CD1* genes, the 3′ UN region was more similar to *mCD1D1* than to *mCD1D2* (Table 2). Supporting this notion, intron 3 of rat CD1 was more similar to *mCD1D1*. Close examination of

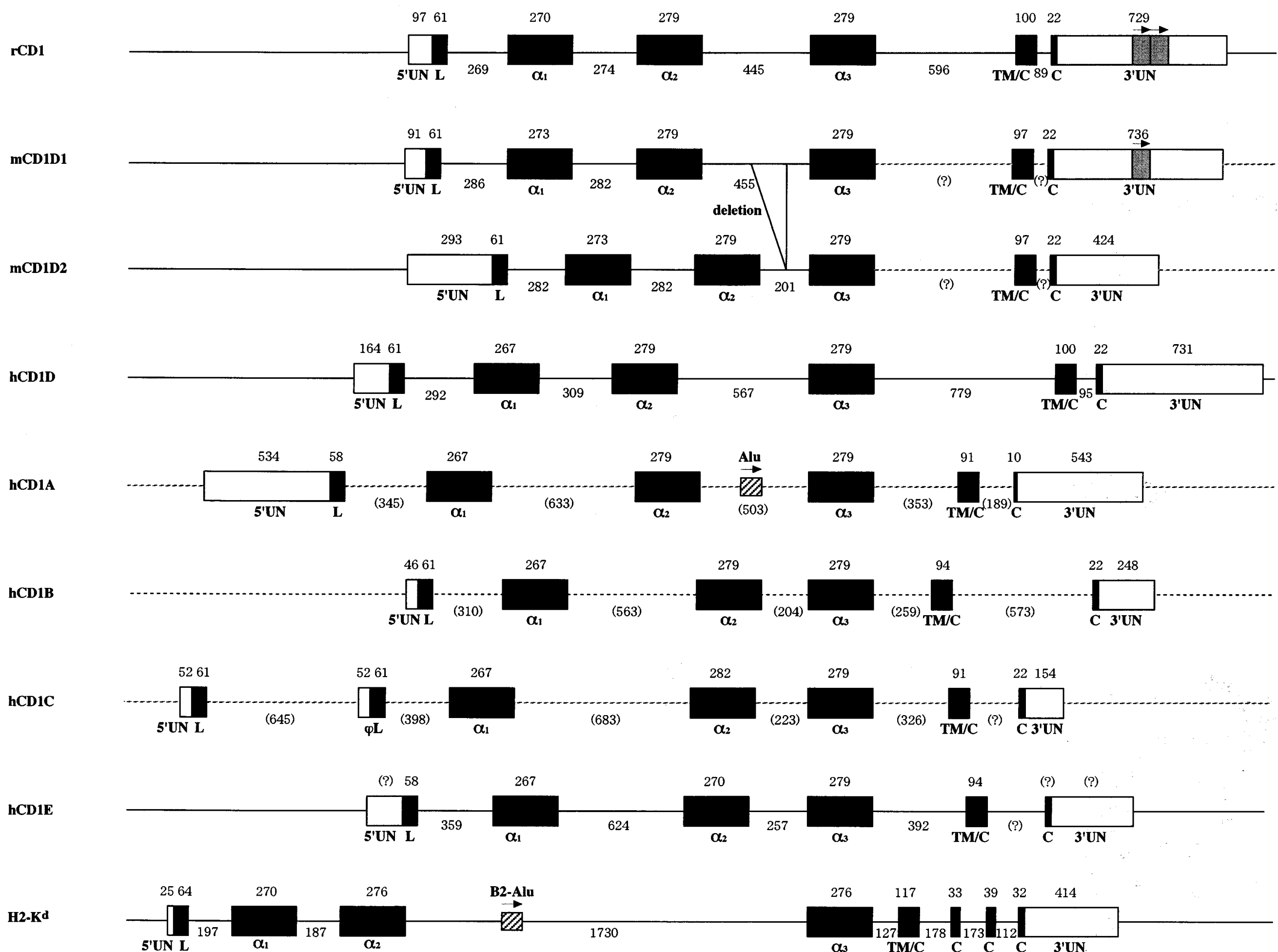


Fig. 4 Comparison of *CD1* gene organization. Exons and introns are indicated by *boxes* and *straight lines*, respectively. *Dotted lines* indicate the sequences not available or not determined. Numbers *above* the boxes are lengths of corresponding exons. Numbers *below* the lines are lengths of corresponding introns. (*?*); region not defined. Numbers in *parentheses*; only length is known and the sequences are not available. References were cited in Table 2. An *Alu*-repeat in *CD1A* is shown by a *hatched box*. DNA stretches duplicated in *rCD1* are also indicated by *arrows*. Organization of *H2-Kd* is shown as a representative of the classical MHC class I gene [Lalanne et al. 1983 (for cDNA) and Kvist et al. 1983 (for gene)]

intron 3 sequences revealed that a 248 bp DNA strech in *mCD1D1* (bases 1308–1555) flanked at both ends by a direct 8 bp repeat (CCTGTGGG) was deleted in *mCD1D2*. An almost identical DNA strech was seen in rat *CD1* but the 8 bp repeat was not clear. This observation indicated that the prototypical rodent *CD1* gene may be *mCD1D1*-like and that the deletion in intron 3 occurring in *mCD1D2* was generated after duplication of the prototype gene in the mouse. Furthermore, in the 3′ UN region of rat *CD1*, there were two highly homologous DNA stretches of 107 bp (1400 to 1506 in 27.1) and 108 bp (1507–1614 in 27.1) which contained a 22 bp dinucleotide (GT) repeat in the 3′ end (Fig. 2). Since this segement had no significant similarity to any other genes except for *mCD1D1*, the duplication of the sequence occurred recently in rat radiation. These findings imply a common origin of rodent *CD1* from an *mCD1D1*-like prototype gene and independent minor alterations in each species.

Comparison of the introns provides an interesting perspective as to the evolution of *CD1* genes found in different species. As shown in Fig. 4 and Table 1, overall organization, including intronic length, could be divided into two types, which were correlated with the previous categorization of two classes of *CD1* genes, classic CD1 and CD1D. For example, the length of intron 1, about 280 bp in *rCD1*, *mCD1s*, and *hCD1D,* was shorter than the approximately 350 bp in classic *CD1* (including *CD1A, CD1B, CD1C, and CD1E*). The length of intron 2 in *CD1D,* about 280 bp, was also shorter than the approximately 600 bp in the classic *CD1*. The lengths of intron 3 and intron 4 of *CD1D* class genes were longer than those of classic *CD1* with only a few exceptions; *mCD1D2* had a deletion in intron 3 and *hCD1A* had an *Alu* repeat in intron 3. Similarly, other intron lengths were also typical of the two classes.

Polymorphism of the rat CD1 locus

In contrast to classical MHC class I genes, there is no evidence for significant polymorphism in the *CD1* gene family by serology, protein, or RFLP analysis (Porcelli

Table 2 Percent similarity of nucleotide sequences of the rat *CD1* gene with those of other species

		rCD1	*mCD1D1*	*mCD1D2*	*hCD1A*	*hCD1B*	*hCD1C*	*hCD1D*	*hCD1E*
5′ un[a]	length[b]	97	91	293	534	46	52	164	ND[e]
	% Similarity[c]		77.6%	77.5%	58.2%	57.5%	54.2%	51.3%	
	(length compared)		(85)	(102)	(55)	(80)	(83)	(78)	
exon 1		61	61	61	58	61	61	61	58
			90.2%	93.4%	52.1%	72.0%	63.8%	65.6%	61.4%
			(61)	(61)	(48)	(25)	(58)	(61)	(57)
intron 1		269	286	282	345	310	645/398	292	359
			72.9%	74.4%	NA[d]	NA	NA	49.5%	49.6%
			(288)	(289				(273)	(274)
exon 2		270	273	273	267	267	267	267	267
			92.9%	91.0%	53.7%	56.3%	57.6%	78.5%	63.7%
			(268)	(268)	(259)	(245)	(255)	(251)	(234)
intron 2		274	282	282	633	563	683	309	624
			82.0%	81.7%	NA	NA	NA	61.1%	49.4%
			(289)	(289)				(311)	(158)
exon 3		279	279	279	279	279	282	279	270
			89.2%	87.5%	59.3%	55.9%	60.1%	75.6%	61.2%
			(279)	(279)	(280)	(281)	(283)	(279)	(281)
intron 3		445	455	201	503	204	223	567	257
			84.5%	66.3%	NA	NA	NA	53.3%	51.0%
			(458)	(172)				(433)	(98)
exon 4		279	279	279	279	279	279	279	279
			91.8%	92.5%	83.8%	83.4%	77.0%	84.9%	83.1%
			(279)	(279)	(278)	(277)	(278)	(278)	(278)
intron 4		596	NA	NA	353	259	326	779	392
			>75	>75	NA	NA	NA	47.2%	55.6%
			85.3%	85.3%				(600)	(117)
			(75)	(75)					
exon 5		100	97	97	91	94	91	100	94
			69.7%	69.7%	50.6%	50.0%	64.4%	56.0%	54.4%
			(99)	(99)	(81)	(68)	(45)	(100)	(68)
intron 5		89	ND[e]	ND	189	573	NA	95	ND
					NA	NA		65.1%	
								(83)	
exon 6		22	22	22	10	22	22	22	ND
			81.8%	81.8%	61.9%	66.7%	86.4%	81.0%	
			(22)	(22)	(21)	(21)	(22)	(21)	
3′ un[a]		729	736	424	543	248	154	731	ND
			80.1%	58.4%	48.6%	45.8%	46.1%	53.8%	
			(734)	(344)	(181)	(216)	(141)	(736)	

[a] The 5′ and 3′ UN regions were not exactly defined and the sequences were based on cDNA clones reported previously
[b] Nucleotide length of exon and intron was shown
[c] Percent similarity to *rCD1* was calculated by GENETYX-Homoam software, version 1.0.2 (GENETYX, SDC, Tokyo) and the length used for comparison with the program is shown in parentheses
[d] NA; Intronic sequences of human *CD1A, CD1B*, and *CD1C* were not available on the public database. Only the length was shown in a paper by Martin and co-workers (1987) Complete intron 4 sequences of mouse *CD1* genes were not available. Only 75 bp sequences adjacent to the exon 4 were reported
[e] ND; Intron 5 sequences of *mCD1* genes were not determined. Since the sequence of the *CD1E* cDNA clone was not reported, parts of *CD1E* (5′ UN region, intron 5, exon 6 and 3′ UN region) could not be determined unambiguously
The sequence sources were as follows: rat *CD1* cDNA (Ichimiya et al. 1994); mouse *CD1* cDNAs and genes (Balk et al. 1991; Bradbury et al. 1988); human *CD1a, CD1b*, and *CD1c* cDNA clones and genes (Aruffo and Seed 1989; Martin et al. 1987); human *CD1d* cDNA (Blumberg et al. 1991); human *CD1D* and *CD1E* genes (Calabi et al. 1989)

Table 3 Allelic variation of rat the *CD1* locus

Inbred rat strains	Codon number
	119
F344/Crj, NIGIII/Hok, LEJ/Hkm, ALB/Hkm, SDJ/Hok, ACI/Hkm, BN/Hok	GCT Ala
LEW/Crj, Wistar/Crj/Smc, TO/Hkm, WKAH/Hkm, W/N/Hkm	GTT Val

1995). We have previously reported that the rat *CD1* locus shows three allelic variations in the 3′ region (Ichimiya et al. 1994). In the present study, the extent of rat CD1 polymorphism was examined for the 5′ region, since recent X-ray crystallographic analysis of the mCD1d1 molecule clarified that a hydrophobic ligand-binding groove is formed by α1 and α2 extracellular domains (Zeng et al. 1997). Southern blot analysis was conducted on eleven laboratory rat strains with a 5′ region probe containing a DNA fragment from exon 1 to exon 3 including introns. Of restriction enzymes so far tested, no RFLP was observed

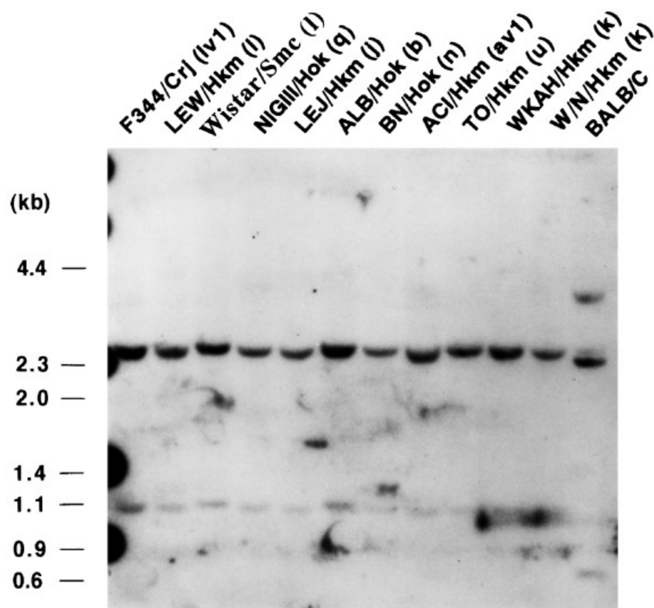


Fig. 5 Southern blot analysis of rat *CD1* locus. Cellular DNAs digested with *Hin*dIII from 11 different rat strains were blotted and probed with a 5' probe containing a DNA fragment from exon 1 to exon 3 (see details in Materials and methods). Two bands (thick 2.4 kb and thin 1 kb) were invariably seen in all rat strains. Slight differences in the mobility reflect the amount of DNA loaded in each lane. Four bands (thick 2.3 kb and 3.5 kb bands, and thin 2.4 kb and 0.6 kb bands) represent two mouse *CD1* genes in BALB/c mice, *mCD1D1,* and *mCD1D2*. The strain name is indicated *above* each lane and the RT1 haplotype is in *parentheses*

with the 5′ probe (Fig. 5). This result indicated that rat CD1 appeared to be less polymorphic in the 5′ region than in the 3′ region.

Polymorphism was further analyzed at the nucleotide level. Exon 2 and exon 3 codings for the α1 and α2 domains, respectively, were amplified from cellular DNA of twelve different rat strains and the sequences were determined including exon-intron boundaries. As shown in Table 3, a single nucleotide substitution (cytosine vs thymine) was found at base 1382 presented in Fig. 2, and caused amino acid alteration at codon 119 of alanine (GCT) to valine (GTT). The alanine was seen in seven strains, including F344/Crj, NIGIII/Hok, LEJ/Hkm, ALB/Hkm, SDJ/Hok, ACI/Hkm, and BN/Hok. The valine was seen in five different strains, LEW/Crj, Wistar/Smc, TO/Hkm, WKAH/Hkm, and W/N/Hkm. Thus, rat CD1 has at least two alleles in the extracellular domains.

Discussion

We determined the complete genomic structure of the rat *CD1* gene, which illuminates the path of *CD1* evolution. Two classes of *CD1* genes, divided by similarity of protein coding regions, could also be typified by their introns. As previously assumed from Southern blot analysis, only a single gene is present in the rat genome. No classic *CD1* class gene was found in the rat. This result confirmed the absence of thymus-specific *CD1* genes (classic *CD1*) in rodents. In contrast, entire regions of *CD1D* have been conserved in rodents and humans. These findings further support a previous hypothesis that classic *CD1* and *CD1D* class genes have different evolutionary histories.

The 5′ flanking regions of rat CD1 contained the binding motifs for transcription factors NF-IL-2A and NF-IL6, whereas some regulatory elements found in MHC class I genes (enhancer A, enhancer B, and IFN response elements) were absent. It is important to determine whether rat CD1 expression is increased in inflamed tissues through these cytokines, as human CD1a-c have been shown to be inducible by IL-4 and GM-CSF on monocytes (Kasinrerk et al. 1993; Porcelli et al. 1992).

Only two allelic variations (alanine vs valine) were detected with sequence analysis of exons encoding the extracellular domains of rat *CD1*. Both amino acids were hydrophobic and the position corresponded to the bottom of the groove (α2 S1 β strand) (Zeng et al. 1997). Whether this allelic dimorphism affects the overall tertiary structure of the groove or the ligand-binding capability, and thus has functional consequences such as in the reactivity of T cells, needs to be clarified. This oligomorphism does not correspond to the MHC haplotypes, indicating that rat *CD1* is not linked to MHC. A few nucleotide changes of the 3′ UN region were found in several rat strains (unpublished observations), which probably corresponded to the results of Southern blot analysis; RFLP was detected with the 3′ region probe but not with the 5′ region probe.

Compared with mouse classical MHC class I genes (Kvist et al. 1983), all *CD1* genes had a quite short intron 3 and long intron 4 (Fig. 4). It is well known that frequent recombinations occur in intron 3 in mouse and rat class I genes (Fisher et al. 1989; Flaherty et al. 1990; Matsuura et al. 1997). Such unequal crossing-over and gene duplication play important roles in creating new members of class I genes (Hughes and Nei 1989). The relatively short intron 3 and absence of highly repetitive elements such as an *Alu*-like sequence within introns of all *CD1s* except for *CD1A* may explain in part why the number of *CD1* family genes is less than that of MHC class I genes.

Although the distribution patterns of mouse TL and CD1 (mCD1d) are different, they were found to be expressed by some of the same cell types such as intestinal epithelium and thymocytes (Bleicher et al. 1990; Brossay et al. 1997; Hershberg et al. 1990; Old et al. 1963; Wu et al. 1991) and share some common properties such as β_2m-dependent, TAP-independent expression (Brutkiewicz et al. 1995; Holcombe et al. 1995; Rogers et al. 1995). It was, therefore, speculated that TL and mCD1d perform similar functions. Two recent reports suggested that NK T cells are numerically increased in TAP-deficient mice and in TL transgenic mice (Joyce et al. 1996) but reduced in CD1-deficient mice (Smiley et al. 1997). TL and CD1 molecules play a role in NK T-cell development in the mouse.

As reported for human CD1d (Blumberg et al. 1991; Canchis et al. 1993), rat CD1 transcripts are expressed by a wide variety of cells and tissues including lymphoid and

nonlymphoid organs (Ichimiya et al. 1994). Rat CD1 expression by intestinal epithelial cells, hepatocytes, renal tubules, and epidermal cells was also shown by in situ hybridization and immunohistochemistry (Burke et al. 1994; Kasai et al. 1997). Thymic expression of rat CD1 appeared to be more prominent than that reported for human CD1d and mCD1s, as we could readily detect its mRNA and protein expression (Ichimiya et al. 1994; Kasai et al. 1997). We also found that rat class Ib genes (*RT1.P*) homologous to mouse *TL* were pseudogenes (Matsuura et al. 1997). Taken together, these findings indicate that rats have a high level of thymic expression of CD1d and lack both classic CD1 and authentic TL antigens. By analogy to mice, functions of classic CD1 and TL may be substituted for in part by CD1d and other class Ib molecules expressed by rat thymus. Rat T cells with an invariant TCR (rat homologue of mouse Vα14-Jα281 and human Va24-JaQ) reacted with CD1-expressing cells (manuscript in preparation). Such unconventional T cells selected by evolutionarily conserved CD1d molecules may interact with a ligand molecule in pathogens and the diet or in cellular components common to mammmals.

Acknowledgments We are grateful to Dr. M. Kinebuchi for helping with the sequence analysis. We also appreciate the financial support provided by The Hokkaido Geriatrics Research Institute.

References

Aizawa, M. and Natori, T. Major histocompatibility complex of the rat, *Rattus norvegicus*. Its structure and function. *Hokkaido University Medical Library Series* (*Vol. 21*), Hokkaido University School of Medicine, Sapporo, Japan, 1988

Aruffo, A. and Seed, B. Expression of cDNA clones encoding the thymocyte antigens CD1a, b, c demonstrates a hierarchy of exclusion in fibroblasts. *J Immunol 143:* 1723–1730, 1989

Balk, S.P., Bleicher, P.A., and Terhorst, C. Isolation and characterization of a cDNA and gene coding for a fourth CD1 molecule. *Proc Natl Acad Sci USA 86*: 252–256, 1989

Balk, S.P., Bleicher, P.A., and Terhorst, C. Isolation and expression of cDNA encoding the murine homologues of CD1. *J Immunol 146:* 768–774, 1991

Bleicher, P.A., Balk, S.P., Hagen, S.J., Blumberg, R.S., Flotte, T.J., and Terhorst, C. Expression of murine CD1 on gastrointestinal epithelium. *Science 250:* 679–682, 1990

Blumberg, R.S., Terhorst, C., P. Bleicher, P.A., McDermott, F.V., Allan, C.H., Landau, S.B., Trier, J.S., and Balk, S.P. Expression of a nonpolymorphic MHC class I-like molecule, CD1D, by human intestinal epithelial cells. *J Immunol 147:* 2518–2524, 1991

Boumsell, L. Cluster report: CD1. *In* W. Knapp (ed.): *Leukocyte Typing IV,* pp. 251–254, Oxford University Press, Oxford, 1989

Bradbury, A., Belt, K.T., Neri, T.M., Milstein, C., and Calabi, F. Mouse CD1 is distinct from and co-exists with TL in the same thymus. *EMBO J 7:* 3081–3086, 1988

Brossay, L., Jullien, D., Cardell, S., Sydora, A.C., Burdin, N., Modlin, R.L., and Kronenberg, M. Mouse CD1 is mainly expressed on hemopoietic-derived cells. *J Immunol 159:* 1216–1224, 1997

Brutkiewicz, R.R., Bennink, J.R., Yewdell, J.W., and Bendelac, A. TAP-independent, β_2-microglobulin-dependent surface expression of functional mouse CD1.1. *J Exp Med 182:* 1913–1919, 1995

Burke, S., Landau, S., Green, R., Tseng, C.C., Nattakom, T., Canchis, W., Yang, L., Kaiserlian, D., Gespach, C., Balk, S.P., and Blumberg, R. Rat cluster of differentiation 1 molecule: expression on the surface of intestinal epithelial cells and hepatocytes. *Gastroenterology 106:* 1143–1149, 1994

Calabi, F. and Milstein, C. A novel family of human major histocompatibility complex-related genes not mapping to chromosome 6. *Nature 323:* 540–543, 1986

Calabi, F., Belt, K.T., Yu, C.-Y., Bradbury, A., Mandy, W.J., and Milstein, C. The rabbit CD1 and the evolutionary conservation of the CD1 gene family. *Immunogenetics 30:* 370–377, 1989a

Calabi, F., Jarvis, J.M., Martin, L., and Milstein, C. Two classes of CD1 genes. *Eur J Immunol 19:* 285–292, 1989b

Calabi, F., Yu, C.Y., Caroline, A.G., Bilsland, C.A., and Milstein, C. CD1: from structure to function. *In* R. Srivastava, B.P. Ram, and P. Tyle (eds.): *Immunogenetics of the Major Histocompatibility Complex*, pp. 215–243, VCH Publishers, New York, 1991

Canchis, P.W., Bhan, A.K., Landau, S.B., Yang, L., Balk, S.P., and Blumberg, R.S. Tissue distribution of the non-polymorphic major histocompatibility complex class I-like molecule, CD1d. *Immunology 80:* 561–565, 1993

Cattoretti, G.E., Berti, E., Parravicini, C., Buscaglia, M., Cappio, F., Caputo, R., Cerri, A., Crosti, L., Dalia, D., Gaiera, G., and Polli, N. Expression of CD1 molecules on dendritic cells: ontogeny, epitope analysis on normal and malignant cells, and tissue distribution. *In* W. Knapp (ed.): *Leukocyte Typing* (*Vol 4*): *White Cell Differentiation Antigens*, pp. 263–264, Oxford University Press, Oxford, 1989

Ferguson, E.D., Dutia, B.M., Hein, W.R., and Hopkins, J. The sheep CD1 gene family contains at least four CD1B homologues. *Immunogenetics 44:* 86–96, 1996

Fisher, D.A., Pecht, M., and Hood, L. DNA sequence of a class I pseudogene from the Tla region of the murine MHC: recombination at a B2 Alu repetitive sequence. *J Mol Evol 28:* 306–312, 1989

Flaherty, L., Elliott, E., Tine, J.A., Walsh, A.C., and Waters, J.B. Immunogenetics of the *Q* and *TL* regions of the mouse. *Crit Rev Immunol 10:* 131–175, 1990

Hershberg, R., Eghtesady, P., Sydora, B., Brorson, K., Cheroutre, H., Modlin, R., and Kronenberg, M. Expression of the thymus leukemia antigen in mouse intestinal epithelium. *Proc Natl Acad Sci USA 87:* 9727–9731, 1990

Holcombe, H.R., Castano, A.R., Cheroutre, H., Teitell, M., Maher, J.K., Peterson, P.A., and Kronenberg, M. Nonclassical behavior of the thymus leukemia antigen: peptide transporter-independent expression of a nonclassical class I molecule. *J Exp Med 181:* 1433–1441, 1995

Hughes, A.L. and Nei, M. Evolution of the major histocompatibility complex: independent origin of nonclassical class I genes in different groups of mammals. *Mol Biol Evol 6:* 559–579, 1989

Ichimiya, S., Kikuchi, K., and Matsuura, A. Structural analysis of the rat homologue of CD1: Evidence for evolutionary conservation of the CD1D class and widespread transcription by rat cells. *J Immunol 153:* 1112–1123, 1994

Itoh, Y., Matsuura, A., Kinebuchi, M., Honda, R., Takayama, S., Ichimiya, S., Kon, S., and Kikuchi, K. Structural analysis of CD3 ζ/η locus of the rat: Expression of ζ but not η transcripts by rat T cells. *J Immunol 151:* 4705–4717, 1993

Joyce, S., Negishi, I., Boesteanu, A., DeSilva, A.D., Sharma, P., Chorney, M.J., Loh, D.Y., and Van Kaer, L. Expansion of natural (NK1+) T cells that express αβ T cell receptors in transporters associated with antigen presentation-1 null and thymus leukemia antigen postitive mice. *J Exp Med 184:* 1579–1584, 1996

Kamps, M.P., Corcoran, L., LeBowitz, J.H., and Baltimore, D. The promoter of the human interleukin-2 gene contains two octamer-binding sites and is partially activated by the expression of Oct-2. *Mol Cel Biol 10:* 5464–5472, 1990

Kasai, K., Matsuura, A., Kikuchi, K., Hashimoto, Y., and Ichimiya, S. Localization of rat CD1 transcripts and protein in rat tissues: an analysis of rat CD1 expression by in situ hybridization and immunohistochemistry. *Clin Exp Immunol 109:* 317–322, 1997

Kasinrerk, W., Baumruker, T., Majdic, O., Knapp, W., and Stockinger, H. CD1 molecule expression on human monocytes induced by granulocyte-macrophage colony-stimulating factor. *J Immunol 150:* 579–586, 1993

Kvist, S., Roberts, L., and Dobberstein, B. Mouse histocompatibility genes: structure and organization of a K^d gene. *EMBO J 2:* 245–254, 1983

Lalanne, J.-L., Delarbre, C., Gachelin, G., Kourilsky, P. A cDNA clone containing the entire coding sequence of a mouse H-2K^d histocompatibility antigen. *Nucleic Acids Res 11:* 1567–1577, 1983

Majello, B., Arcone, R., Toniatti, C., and Ciliberto, G. Constitutive and IL-6-induced nuclear factors that interact with the human C-reactive protein promoter. *EMBO J 9:* 457–465, 1990

Martin, L.H., Calabi, F., and Milstein, C. Isolation of CD1 genes: a family of major histocompatibility related differentiation antigens. *Proc Natl Acad Sci USA 83:* 9154–9158, 1986

Matsuura, A., Takayama, S., Kinebuchi, M., Hashimoto, Y., Kasai, K., Kozutsumi, D., Ichimiya, S., Honda, R., Natori, T., and Kikuchi, K. RT1.P, rat class Ib genes related to mouse TL: evidence that CD1 molecules but not authentic TL antigens are expressed by rat thymus. *Immunogenetics 46:* 293–306, 1997

McMichael, A.J., Pilch, J.R., Galfre, G., Mason, D.Y., Fabre, J.W., and Milstein, C. A human thymocyte antigen defined by a hybrid myeloma monoclonal antibody. *Eur J Immunol 9:* 205–210, 1979

Old, L.J., Boyse, E.A., and Stockert, E. Antigenic properties of experimental leukemias. I. Serological studies in vitro with spontaneous and radiation-induced leukemias. *JNCI 31:* 977–986, 1963

Porcelli, S.A., Morita, C.T., and Brenner, M.B. CD1b restricts the response of human CD4$^-$CD8$^-$ cytolytic T lymphocytes to a microbial antigen. *Nature 360:* 593–597, 1992

Porcelli, S.A. The CD1 family: A third lineage of antigen-presenting molecules. *Adv Immunol 59:* 1–98, 1995

Rogers, J.R., Metha, V., and Cook, R.G. Surface expression of β_2-microglobulin-associated thymus-leukemia antigen is independent of TAP2. *Eur J Immunol 25:* 1001–1007, 1995

Sambrook, J., Fritsch, E.F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, 1989

Smiley, S.T., Kaplan, M.H., and Grusby, M. Immunoglobulin E production in the absence of interleukin-4-secreting CD1-dependent cells. *Science 275:* 977–979, 1997

Sugita, M., Jackman, R.M., van Donselaar, E., Behar, S.M., Rogers, R.A., Peters, P.J., Brenner, M.B., and Porcelli, S.A. Cytoplasmic tail-dependent localization of CD1b antigen-presenting molecules to MIICs. *Science 273:* 349–352, 1996

Teunissen, M.B.M., Wormmeester, J., and Kreig, S.R., Peters, P.J., Vogels, I.M.C., Kapsenberg, M.L., and Bos, J.D. Human epidermal Langerhans cells undergo profound morphologic and phenotypical changes during in vitro culture. *J Invest Dermatol 94:* 166–173, 1990

Ting, J.P.-Y. and Baldwin, A.S. Regulation of MHC gene expression. *Curr Opin Immunol 5:* 8–12, 1993

Wu, M., Van Kaer, L., Itohara, S., and Tonegawa, S. Highly restricted expression of the thymus leukemia antigens on intestinal epithelial cells. *J Exp Med 174:* 213–217, 1991

Zeng, Z.-H., Castano, A.R., Segelke, B.W., Stura, E.A., Peterson, P.A., and Wilson, I.A. Crystal structure of mouse CD1: an MHC-like fold with a large hydrophobic binding groove. *Science 277:* 339–345, 1997