Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 252–256, January 1989
Immunology

Isolation and characterization of a cDNA and gene coding for a fourth CD1 molecule

(thymocytes/Langerhans cells/major histocompatibility complex/β_2-microglobulin)

STEVEN P. BALK*†‡, PAUL A. BLEICHER*§, AND COX TERHORST*¶

*Laboratory of Molecular Immunology, Dana–Farber Cancer Institute, Boston, MA 02115; †Department of Hematology and Oncology, Beth Israel Hospital, Boston, MA 02115; and §Department of Dermatology and ¶Department of Pathology, Harvard Medical School, Boston, MA 02115

Communicated by Lloyd J. Old, September 16, 1988 (received for review May 10, 1988)

ABSTRACT **The CD1 locus encodes a family of major histocompatibility complex (MHC) antigen-like glycoproteins which associate with β_2-microglobulin and are expressed on immature thymocytes and Langerhans cells. Three CD1 molecules have been identified by monoclonal antibodies and molecular cloning: CD1a, -b, and -c. We have isolated a cDNA coding for a fourth CD1 molecule from a human thymocyte library and termed this molecule CD1d. Reported here are the complete nucleotide sequence and genomic organization of CD1d. They predict that this molecule is related to the previously identified CD1a, -b, and -c molecules and to MHC class I molecules, with three external domains, a transmembrane domain, and a short cytoplasmic tail. The sequence of CD1d is the most divergent among the CD1 molecules in the membrane-distal α1 and α2 domains and in the 5′ untranslated region. In contrast, all four CD1 molecules are highly homologous in the membrane-proximal α3 domain, which is likely involved in β_2-microglobulin binding. A comparison of CD1 and MHC class I sequences suggests that these molecules each evolved to interact with a distinct set of cell surface proteins.**

The CD1 molecules are a family of major histocompatibility complex (MHC) class I antigen-like glycoproteins expressed predominantly on immature cortical thymocytes (1–3) and originally considered as possible human homologues of murine TL antigens (4). Monoclonal antibodies generated against human thymocytes have identified three CD1 molecules: CD1a, -b, and -c (5–7). All three consist of a glycosylated heavy chain of 43–49 kDa associated with β_2-microglobulin (8–12). The CD1a molecule, defined by the T6 monoclonal antibody, also appears to associate with CD8 (T8) on the thymocyte cell surface (13, 14). Unlike class I MHC molecules, there is no evidence that CD1 molecules are polymorphic (15), and the tissue distribution of CD1 is highly restricted. In addition to expression on immature thymocytes, CD1 expression has been found on some lymphoid malignancies (7) and Langerhans cells in the skin (16), and CD1c has been found on a subpopulation of B cells (17).

A CD1a cDNA was cloned previously with the aid of oligonucleotide probes (18), and cDNAs encoding all three CD1 molecules have been isolated (A. Aruffo and B. Seed, personal communication) by utilizing a high-level expression system in COS cells (19). Genomic clones for these molecules have also been isolated, and their intron/exon organization suggests a domain structure analogous to MHC class I antigens with three extracellular domains (α1, α2, and α3), a transmembrane domain, and a short cytoplasmic tail (15). There is, however, very limited sequence homology between CD1 and MHC molecules (15, 20), including murine TL antigens. Moreover, the CD1 gene is not MHC linked, as it has been mapped to chromosome 1 (21).

We report here the complete nucleotide sequence‖ and genomic organization of a CD1 molecule, CD1d, that has not been detected previously by monoclonal antibodies. The α3 domain of CD1d is highly homologous to that of CD1a, -b, and -c and corresponds exactly to a genomic α3 domain isolated previously (20). In contrast, there is considerable divergence in the α1, α2, and 5′ untranslated regions between CD1d and CD1a, -b, and -c. Sequence comparisons between CD1 and MHC molecules and their possible structural and functional implications are discussed.

MATERIALS AND METHODS

cDNA Probes. Full-length single-stranded CD1a, -b, and -c cDNAs (A. Aruffo and B. Seed, personal communication) were radiolabeled by primer extension with a 3′ hybridization primer (22). Double-stranded cDNA inserts were isolated in low-melting-point agarose and labeled by random oligonucleotide priming (23). The 5′-specific CD1d probe extended from the 5′ *Eco*RI linker to an internal EcoRI site at nucleotide 465. The 3′-specific probe extended from a *Pst* I site at nucleotide 1548 to the 3′ *Eco*RI linker.

Library Screening. The human thymus cDNA library in λgt11 was obtained from Clontech Laboratories (Palo Alto, CA); it contained ≈2 × 10^6 recombinants and was amplified once according to the supplier. Hybridizations were carried out with a mixture of CD1a, -b, and -c probes in a buffer containing 50% (vol/vol) formamide and 5× SSC at 37°C and washed in 0.1× SSC/0.1% sodium dodecyl sulfate at 50°C according to standard methods (ref. 24, pp. 326–328; 1× SSC is 0.15 M NaCl/0.015 M sodium citrate, pH 7.0). The EMBL3 human genomic library was a partial *Mbo* I digest of genomic DNA (kindly provided by Stuart Orkin, Children's Hospital, Boston) and was screened with a mixture of CD1a, -b, and -c probes as above.

DNA Sequencing. A full-length CD1d cDNA was produced by partial *Eco*RI digestion, isolated, and subcloned in pGEM-4 (Promega Biotec, Madison, WI). Constructs were then made in M13 or pGEM-4 and were sequenced by the dideoxy chain termination method (25), using Sequenase (United States Biochemical, Cleveland) or the Klenow fragment of DNA polymerase. The majority of the cDNA was sequenced from both strands and the sequences around all restriction sites used for constructs were confirmed on overlapping constructs. Genomic exon containing fragments were subcloned into pGEM-4 and partially sequenced as above.

RNA Analysis. RNA from neonatal human thymus was extracted in LiCl/urea (26), separated on formaldehyde/agarose gels (ref. 24, p. 202), and transferred to nylon membranes (Biotrans, ICN). The blots were hybridized in

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.

Abbreviation: MHC, major histocompatibility complex.

‡To whom reprint requests should be addressed.

‖The sequence reported in this paper is being deposited in the EMBL/GenBank data base (accession no. J04142).

buffer containing 50% formamide and 5× SSC at 42°C and washed in 0.2× SSC/0.1% sodium dodecyl sulfate at 65°C.

RESULTS

Isolation of CD1 cDNAs. Approximately 2×10^5 plaques of a human cDNA library in bacteriophage λgt11 were screened with a mixture of full-length CD1a, -b, and -c probes. Six independent bacteriophage isolates were purified and found to contain identically sized 1.9-kilobase (kb) inserts with an internal *Eco*RI site yielding fragments 0.5 and 1.4 kb long. These fragments from one isolate (λgt11ht-5) were subcloned and partially sequenced. This partial sequence indicated that bacteriophage λgt11ht-5 carried cDNA related to but distinct from CD1a, -b, and -c cDNAs. Moreover, the 0.5-kb *Eco*RI fragment corresponded to the 5′ end of the insert and did not extend to the highly homologous α3 domain. A probe made from the 0.5-kb *Eco*RI fragment hybridized specifically to the 0.5-kb fragment of all six isolates but not to CD1a, -b, or -c cDNA (data not shown), further indicating that all six isolates were identical. The bacteriophage λgt11ht-5 cDNA was therefore further subcloned and sequenced.

Nucleotide and Deduced Amino Acid Sequence of λgt11ht-5. The complete nucleotide sequence and the deduced amino acid sequence of λgt11ht-5 are shown in Fig. 1. The first ATG occurred at position 165 and was in good agreement with the eukaryotic consensus initiation site (purine at −3 and G at +4) (27). This ATG was followed by a long open reading frame which ended at 1170, encoding a protein of 335 amino acids. Nucleotides 772–1050 corresponded exactly to the α3 domain exon sequenced previously from a genomic clone (20). The four potential sites for N-linked glycosylation are indicated in Fig. 1. The unglycosylated translation product had a predicted molecular weight of 37,674 (35,459 without the leader peptide). There was a long 3′ untranslated region of 705 base pairs (bp) with no poly(A) tail or polyadenylylation signal. Sequencing of the corresponding genomic region (see below) revealed a probable polyadenylylation signal 21 bp from the end of this cDNA (underlined in Fig. 1).

Comparison Between the Amino Acid Sequences of λgt11ht-5 and CD1a, -b, and -c. Comparison of the amino acid sequences predicted from the λgt11ht-5 cDNA and the CD1a, -b, and -c cDNAs (ref. 15 and A. Aruffo and B. Seed, personal communication) indicated homologous molecules with similar domain structures (Fig. 2). These similar domain structures were confirmed by the genomic organization (see below). The most homologous region between all four molecules was the putative β_2-microglobulin-binding α3 domain, with >90% identity (20). On the basis of the domain structure and sequence homology, we termed the protein encoded by the bacteriophage λgt11ht-5 cDNA CD1d.

In contrast to the α3 domain, the α1 and α2 domains of CD1d and CD1a, -b, and -c diverged considerably (Fig. 2). The low homology in these domains between CD1a, -b, and -c has been noted previously (15), with 48–56% of residues conserved between CD1a, -b, and -c. CD1d has diverged somewhat further, with 37–40% of residues conserved between CD1d and CD1a, -b, and -c in the α1 and α2 domains (Fig. 2).

Four potential sites for N-glycosylation were present in each of the CD1 molecules. Three of these sites were conserved between CD1a, -b, and -c, while only one site (Asn-18) was conserved in all four molecules. Invariant Cys residues were present at positions 100 and 164 in the α2 domain which, based on homology with HLA class I molecules (see below), are probably linked by an intrachain disulfide bridge. A third Cys, which was conserved in the α2 domain of CD1a, -b, and -c at position 143, was an Ile in CD1d. A Pro-Xaa-Pro sequence in the α2 domain at position 131 was also conserved in CD1a, -b, and -c (15) but in CD1d was a Pro-Xaa-Gln.

Genomic Organization of CD1d. A human genomic DNA library in the EMBL3 vector was screened with a mixture of CD1 probes. Positive bacteriophage were rescreened with probes corresponding to the CD1d 5′ untranslated exon through the α1 exon and the 3′ untranslated region. Several overlapping bacteriophage were isolated, one of which (EMBL3-hg14) appeared to contain the entire gene. This bacteriophage was mapped and subcloned and exon containing regions were sequenced.

The intron/exon organization of the CD1d gene is shown in Fig. 3 and sequences around the intron/exon boundaries

```
cagaagcagcaaaccgccggcaagcccagcgaggagggctgccggggtctgggcttgggaattggctggcacccagcggaaagggacgtgagctgagcgcgggggg  105
                                                            ■L →
agaagagtgcgcaggtcagagggcggcgcgcagtccgcgaggtccccacgccgggcgat ATG GGG TGC CTG CTG TTT CTG CTG CTC TGG GCG CTC  200
                                                            Met Gly Cys Leu Leu Phe Leu Leu Leu Trp Ala Leu   -9
                                ■α1 →                                                                *
CTC CAG GCT TGG GGA AGC GCT GAA GTC CCG CAA AGG CTT TTC CCC CTC CGC TGC CTC CAG ATC TCG TCC TTC GCC AAT AGC  281
Leu Gln Ala Trp Gly Ser Ala Glu Val Pro Gln Arg Leu Phe Pro Leu Arg Cys Leu Gln Ile Ser Ser Phe Ala Asn Ser   19
                                                                                 *
AGC TGG ACG CGC ACC GAC GGC TTG GCG TGG CTG GGG GAG CTG CAG ACG CAC AGC TGG AGC AAC GAC TCG GAC ACC GTC CGC  362
Ser Trp Thr Arg Thr Asp Gly Leu Ala Trp Leu Gly Glu Leu Gln Thr His Ser Trp Ser Asn Asp Ser Asp Thr Val Arg   46

TCT CTG AAG CCT TGG TCC CAG GGC ACG TTC AGC GAC CAG CAG TGG GAG ACG CTG CAG CAT ATA TTT CGG GTT TAT CGA AGC  443
Ser Leu Lys Pro Trp Ser Gln Gly Thr Phe Ser Asp Gln Gln Trp Glu Thr Leu Gln His Ile Phe Arg Val Tyr Arg Ser   73
                                                                ■α2 →
AGC TTC ACC AGG GAC GTG AAG GAA TTC GCC AAA ATG CTA CGC TTA TCC TAT CCC TTG GAG CTC CAG GTG TCC GCT GGC TGT  524
Ser Phe Thr Arg Asp Val Lys Glu Phe Ala Lys Met Leu Arg Leu Ser Tyr Pro Leu Glu Leu Gln Val Ser Ala Gly Cys  100
                     *
GAG GTG CAC CCT GGG AAC GCC TCA AAT AAC TTC TTC CAT GTA GCA TTT CAA GGA AAA GAT ATC CTG AGT TTC CAA GGA ACT  605
Glu Val His Pro Gly Asn Ala Ser Asn Asn Phe Phe His Val Ala Phe Gln Gly Lys Asp Ile Leu Ser Phe Gln Gly Thr  127

TCT TGG GAG CCA ACC CAA GAG GCC CCA CTT TGG GTA AAC TTG GCC ATT CAA GTG CTC AAC CAG GAC AAG TGG ACG AGG GAA  686
Ser Trp Glu Pro Thr Gln Glu Ala Pro Leu Trp Val Asn Leu Ala Ile Gln Val Leu Asn Gln Asp Lys Trp Thr Arg Glu  154
                         *
ACA GTG CAG TGG CTC CTT AAT GGC ACC TGC CCC CAA TTT GTC AGT GGC CTC CTT GAG TCA GGG AAG TCG GAA CTG AAG AAG  767
Thr Val Gln Trp Leu Leu Asn Gly Thr Cys Pro Gln Phe Val Ser Gly Leu Leu Glu Ser Gly Lys Ser Glu Leu Lys Lys  181
    ■α3 →
CAA GTG AAG CCC AAG GCC TGG CTG TCC CGT GGC CCC AGT CCT GGC CCT GGC CGT CTG CTG CTG GTG TGC CAT GTC TCA GGA  848
Gln Val Lys Pro Lys Ala Trp Leu Ser Arg Gly Pro Ser Pro Gly Pro Gly Arg Leu Leu Leu Val Cys His Val Ser Gly  208

TTC TAC CCA AAG CCT GTA TGG GTG AAG TGG ATG CGG GGT GAG CAG GAG CAG CAG GGC ACT CAG CCA GGG GAC ATC CTG CCC  929
Phe Tyr Pro Lys Pro Val Trp Val Lys Trp Met Arg Gly Glu Gln Glu Gln Gln Gly Thr Gln Pro Gly Asp Ile Leu Pro  235

AAT GCT GAC GAG ACA TGG TAT CTC CGA GCA ACC CTG GAT GTG GTG GCT GGG GAG GCA GCT GGC CTG TCC TGT CGG GTG AAG 1010
Asn Ala Asp Glu Thr Trp Tyr Leu Arg Ala Thr Leu Asp Val Val Ala Gly Glu Ala Ala Gly Leu Ser Cys Arg Val Lys  262
                                                    ■TM →
CAC AGC AGT CTA GAG GGC CAG GAC ATC GTC CTC TAC TGG GGT GGG AGC TAC ACC TCC ATG GGC TTG ATT GCC TTG GCA GTC 1091
His Ser Ser Leu Glu Gly Gln Asp Ile Val Leu Tyr Trp Gly Gly Ser Tyr Thr Ser Met Gly Leu Ile Ala Leu Ala Val  289
                                                        ■C →
CTG GCG TGC TTG CTG TTC CTC CTC ATT GTG GGC TTT ACC TCC CGG TTT AAG AGG CAA ACT TCC TAT CAG GGC GTC CTG tga 1172
Leu Ala Cys Leu Leu Phe Leu Leu Ile Val Gly Phe Thr Ser Arg Phe Lys Arg Gln Thr Ser Tyr Gln Gly Val Leu      315

ctcgccttgccacatctgtgtctctggaacccaggacctctggacctcaggttcccaagacttcagtcctggtctgctcaggaattgaagatgtaaggaattgaaga 1279
taggagagataccttgaaaaagtagagaacagtcatgaggcagctttcatcacaccctttttaacatttatctaaaagaatttaaattctttttcaaaaattacacta 1386
caagtttataagcccaaatggctctgtgaaatcagaagtgcaaaggtgtgcaaacttgtatctgaagacctaccagggacaagcaggtaagagctgatgtgagtgtg 1493
tgtgatgggatctgtaaggaactggaacacacatgtcctatccaaaggaatcagctgcagctgcttgttgtcaagtataaagtcaggacctggcttggctttaaccg 1600
tttttcaagaaaactggaaatctggattttcagcgaacatgcctgattttaaaaggttgactcaagtttttacaaaatactatgtgggacacctcaaatacatacct 1707
actgactgatgacaaacccaggagtttgtgtgtcttttataaaaagtttgccctggatgtcatattggcagttggaggacacagtttctattgtaaatttggattta 1814
cgactgaagaaggacattttctctttaaaagaaagttaggttataagaaacagaggcgtctcacatttttacttggtgtaattaataaa                  1903
```

FIG. 1. Complete nucleotide sequence and deduced amino acid sequence of λgt11ht-5 cDNA. The cDNA ends at nucleotide 1877; the underlined sequence from 1878 to the polyadenylylation signal was deduced from genomic sequencing. The positions of the leader (L), extracellular (α1–α3), transmembrane (TM), and cytoplasmic (C) regions are indicated, and numbering of the protein begins with the first amino acid after the leader. Potential sites for N-glycosylation are indicated with asterisks.

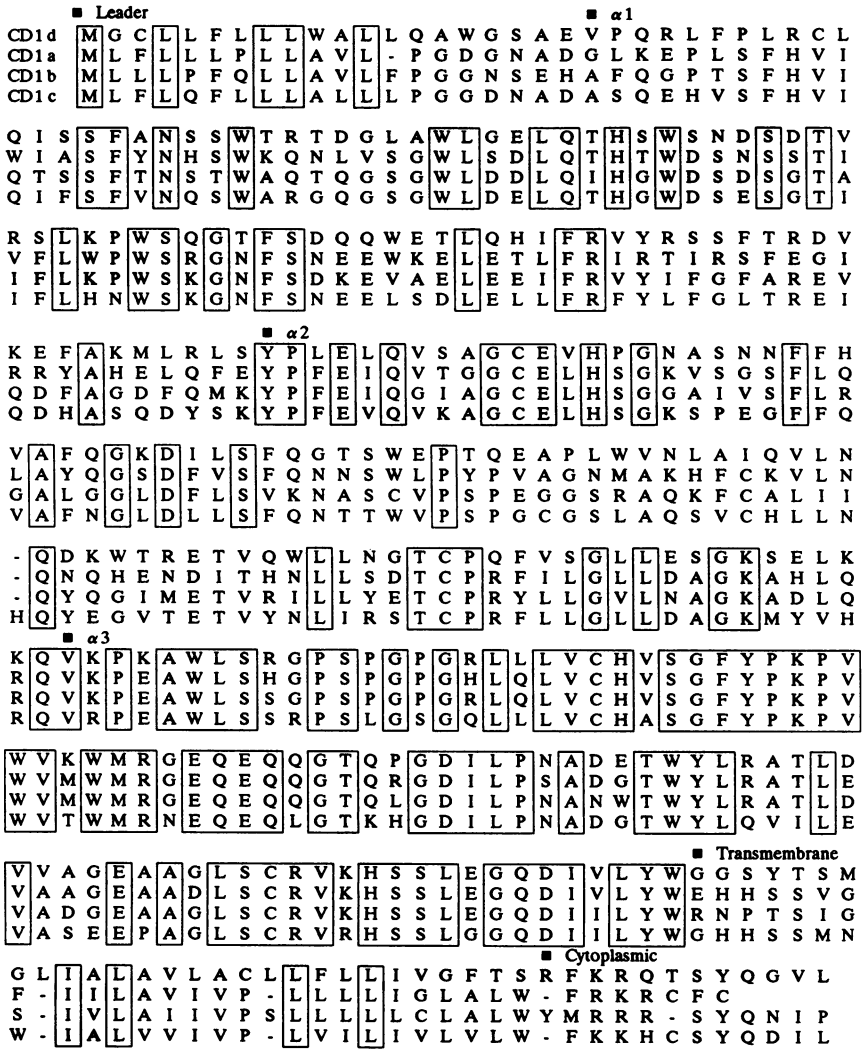

FIG. 2. Amino acid homology among CD1 molecules. Amino acid sequences for CD1a, -b, and -c were deduced from their cDNAs (A. Aruffo and B. Seed, personal communication). The domain structure was based on homology with MHC class I molecules (28) and on the exon organization (except the first basic residue after the hydrophobic transmembrane region is indicated as the start of the cytoplasmic tail). Residues conserved in all four molecules are boxed.

are given in Table 1. Exon 1 encoded the 5′ untranslated and leader region. Exons 2, 3, and 4 encoded the extracellular α1, α2, and α3 domains, respectively. The transmembrane region, a short cytoplasmic tail, and 3′ untranslated region were on exons 5 and 6. Exon 6 most likely extended through the polyadenylylation signal identified from the genomic sequence (Fig. 1). The genomic organization of CD1d was therefore identical to that of CD1a, -b, and -c (15).

Northern Blot Analysis of CD1d. The isolation of multiple CD1d cDNAs from a thymocyte library suggested that CD1d might be a high-frequency transcript in thymus. To determine whether CD1d was a relatively abundant transcript, Northern blots of thymus RNA were hybridized with full-length CD1a and CD1d probes. Both probes hybridized to transcripts of about 2 kb, with significantly less hybridization of the CD1d probe (Fig. 4). The full-length CD1 probes may cross-hybridize with other CD1 messages via their conserved α3 domain. Hybridization of thymus RNA with a probe specific for the 3′ untranslated region of CD1d could be detected only with a long exposure (data not shown). These results indicated that CD1d is a relatively low-abundance transcript in thymus.

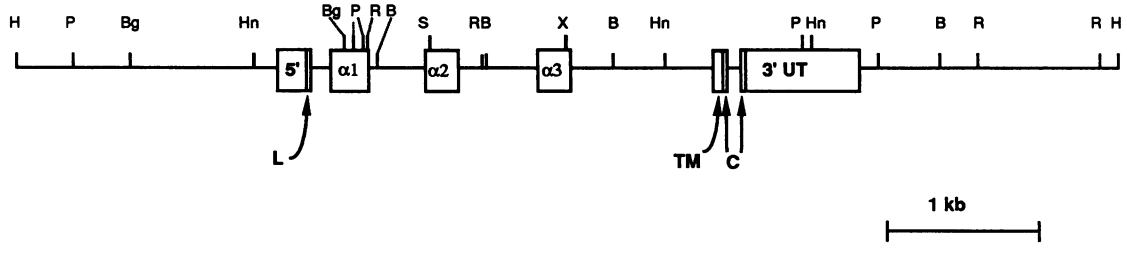

FIG. 3. Restriction map and intron/exon organization of the CD1d gene. Exons (boxes): 5′, 5′ untranslated region; L, leader; α1–α3, extracellular α1, α2, and α3 domains; TM, transmembrane region; C, cytoplasmic tail; and 3′ UT, 3′ untranslated region. Restriction sites in the EMBL3hg-14 CD1d genomic phage are indicated: B, BamHI; Bg, *Bgl* II; R, *Eco*RI; H, *Hin*dIII; Hn, *Hin*cII; P, *Pst* I; and X, *Xba* I.

Table 1. Intron/exon boundaries of CD1d

Intron	Boundary	Exon	Boundary	Intron
		5' UT	---Ala-Glu-V ---GCT GAA G	gtggtggaa
ctctcacag	al-Pro-Gln--- TC CCG CAA---	α1	---Leu-Ser-T ---TTA TCC T	gtgagctga
tctccacag	yr-Pro-Leu--- AT CCC TTG---	α2	---Lys-Gln-V ---AAG CAA G	gtcagcctg
tggatacag	al-Lys-Pro--- TG AAG CCC---	α3	---Tyr-Trp-G ---TAC TGG G	gtgagaaaa
ccagagcag	ly-Gly-Ser--- GT GGG AGC---	TM/C	---Arg-Gln-Th ---AGG CAA AC	gtaagtctc
ctctcacag	r-Ser-Tyr--- T TCC TAT---	C/3' UT		

Boundary nucleotides for each intron/exon boundary are shown. Introns are in lowercase letters, exons in capitals. The corresponding amino acid sequence is listed above each exon. Exon abbreviations are as in Fig. 3 legend.

Comparisons Between Amino Acid Sequences of CD1 and MHC Molecules. Comparison of the CD1 predicted amino acid sequence with HLA class I and II molecule sequences (29) revealed several noteworthy features. There was very little homology between the CD1 and HLA class I α1 domains, with an optimal alignment yielding at most seven potentially conserved residues (data not shown). In contrast, the α2 domains of these molecules could be aligned to yield conserved consensus residues at 24 positions (26%; Fig. 5*A*). The presence of several additional conservative changes and identity between at least one CD1 residue and HLA at variable positions (not shown but compare Figs. 2 and 5*A*) further indicated that these α2 domains were related.

The low homology between class I and CD1 α3 domains and the comparable homology between class II β_2 and CD1 α3 domains have been noted (20). Alignment of consensus sequences for CD1a–d indicated there was significantly less homology between CD1 and HLA class I α3 domains (31% of residues conserved) than between CD1α3 and HLA class II β_2 domains (42% of residues conserved; Fig. 5*B*). Also noteworthy was that the possible contact points between the HLA class I α3 domain and β_2-microglobulin (28) were not conserved in CD1 (Fig. 5*B*, indicated in black squares), except for an Asp at position 238 in HLA.

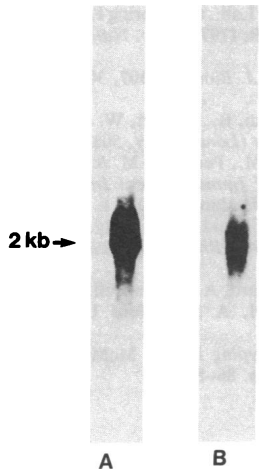

FIG. 4. Northern blot of thymus RNA. RNA was extracted from neonatal human thymus and 20 μg was electrophoresed, transferred, and hybridized. Duplicate blots were hybridized with a full-length CD1a (*A*) or CD1d (*B*) probe and exposed for 16 hr.

DISCUSSION

A cDNA encoding an additional member of the CD1 gene family has been isolated from a human thymus cDNA library and termed CD1d. This cDNA encodes a polypeptide of 335 amino acids with significant homology to CD1a, -b, and -c (Figs. 1 and 2). The overall structure and genomic organization, with three external domains, a transmembrane region, and a short cytoplasmic tail, are identical in all four CD1 molecules (Fig. 3). The sequence homology is greatest in the putative β_2-microglobulin-binding α3 domain, which is highly conserved among all four molecules (Fig. 2). In contrast, the CD1d α1 and α2 domains have diverged considerably. CD1d has the least homology among the CD1 molecules and has no significant homology with CD1a, -b, or -c in the 5′ untranslated region (Fig. 1 and ref. 15). The significance of homology in the 5′ untranslated region between CD1a, -b, and -c is unclear, but it could reflect one level of posttranscriptional control that is not shared by CD1d.

Northern blot analysis indicates that CD1d message is present in low abundance in thymus (Fig. 4). The relatively high frequency of CD1d isolates from the thymus library used here probably reflects amplification of this cDNA, but variability in CD1d expression cannot be ruled out. We have noted only a low expression of CD1d in two separate thymus RNA preparations and low or undetectable expression in several T-lymphoblastoid cell lines (data not shown). A non-full-length cDNA which may correspond to CD1d has been found in a subclone of MOLT-4 selected for high CD1a expression (20). No detectable message has been found in a series of B-cell or myeloid lines (data not shown). Attempts to identify a fourth CD1 molecule serologically have been suggestive but indirect (5, 7). Whether CD1d message is present uniformly in thymus at low levels or is expressed by only a small subset of thymus cells cannot yet be determined.

The structure and distribution of CD1 suggest that these molecules function by interaction with T-cell receptors. In analogy with MHC class I and II molecules, CD1 on Langerhans cells or on some thymocytes may bind and present antigen to mature T lymphocytes bearing a particular class of T-cell receptors. A second model involves CD1 interaction with newly rearranged T-cell receptor chains on the surface of immature thymocytes as an initial restriction element. In this general model, an appropriate interaction between CD1 and TCR would signal the immature thymocyte to express that TCR complex and proceed with further maturation and selection events in the thymus.

There has been considerable divergence between CD1 and MHC class I molecules despite their similar domain structure and binding of β_2-microglobulin. The α1 domain is most divergent between CDl and MHC class I molecules, suggesting that there are different structural constraints on the α1 and α2 domains. Alignment of the α2 domains (Fig. 5) reveals some noteworthy features. The Pro-Xaa-Pro at position 131 in CD1a, -b, and -c aligns with a non-α-helical area in HLA class I (28). The invariant cysteines in CD1 at positions 101 and 164 align with invariant disulfide-linked cysteines in MHC class I molecules (28). There appears to be some pattern to the conserved residues around Cys-164, with a conserved residue in every third or fourth position. If this region forms an α-helix similar to HLA class I (28), then one face of the helix could be highly conserved. This region of class I (Cys-164 to Gln-180) is also closely associated with the α1 domain in HLA (28) and could be similarly involved with CD1 intra- or intermolecular associations.

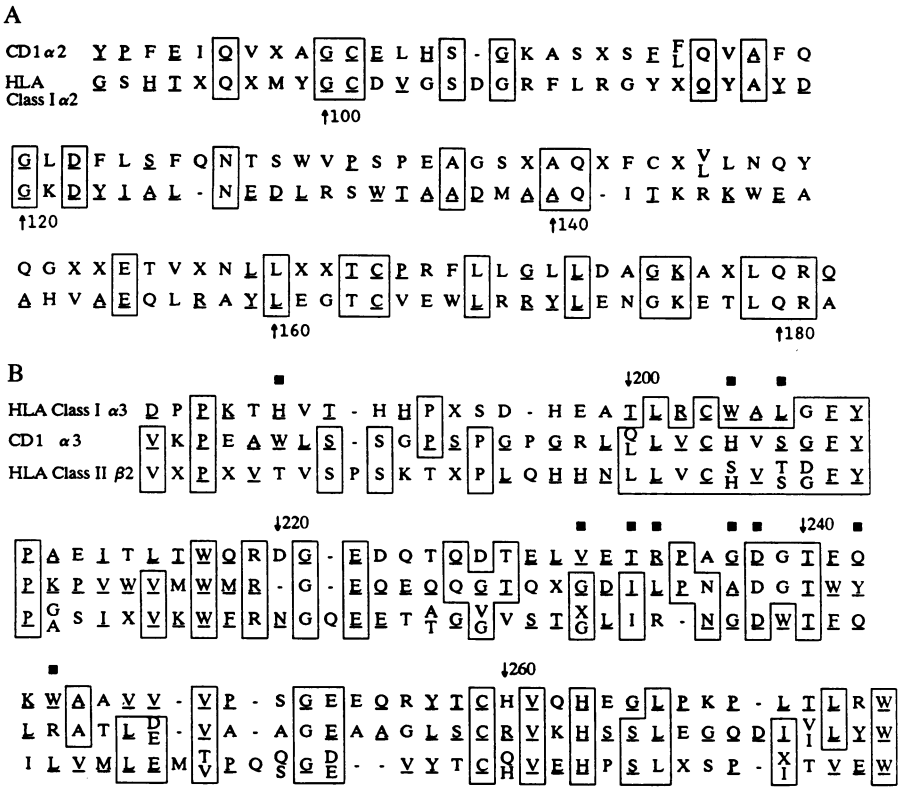

FIG. 5. Homology between CD1 and MHC molecules. Consensus sequences (50% or greater) for CD1 and human MHC molecules (29) were aligned, and residues conserved between CD1 and MHC are boxed. Residues which are invariant in CD1, MHC class I, or class II (present in all human and nonhuman isolates) are underlined. The numbering is based on HLA class I residues. (*A*) Alignment of CD1 and HLA class I $\alpha 2$ domains. (*B*) Alignment of CD1 and HLA class I $\alpha 3$ domains and the HLA class II $\beta 2$ domain. Residues which are possible β_2-microglobulin contact points (28) in the HLA class I $\alpha 3$ domain are indicated by black squares.

Most striking is how far the $\alpha 3$ domains of CD1 and MHC class I molecules have diverged, as this region is involved in β_2-microglobulin binding and is conserved among MHC class I molecules (Fig. 5*B*). Moreover, the $\alpha 3$ domains of CD1a, -b, -c, and -d are also highly conserved. This suggests the presence of unique constraints on the CD1 $\alpha 3$ domain other than binding of β_2-microglobulin. Whether these constraints involve intramolecular interactions not present in MHC class I molecules or involve CD1 associations with molecules other than β_2-microglobulin remains to be determined. In addition to an association between CD1 and CD8 (13, 14), there is evidence suggesting that other cell surface molecules may interact with CD1 (30–32). These associations may be important clues in understanding the function of this gene family.

We thank Drs. A. Aruffo and B. Seed for providing the CD1a, -b, and -c cDNAs and for helpful discussions. We also thank Mr. John W. Lockhart for help with the figures. This work was supported by Grant IM-289 from the American Cancer Society. S.P.B. was supported by Clinical Investigator Award CA01310 and P.A.B. by Grant AR07098, both from the National Institutes of Health.

1. McMichael, A. J., Pilch, J. R., Galfre, G., Mason, D. Y., Fabre, J. W. & Milstein, C. (1979) *Eur. J. Immunol.* **9**, 205–210.
2. Reinherz, E. L., Kung, P. C., Goldstein, G., Levey, R. H. & Schlossman, S. F. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1588–1592.
3. Knowles, R. K. & Bodmer, W. F. (1982) *Eur. J. Immunol.* **12**, 676–681.
4. Old, L. J. & Stockert, E. (1977) *Annu. Rev. Genet.* **11**, 127–160.
5. Amiot, M., Bernard, A., Raynal, B., Knapp, W., Deschildre, C. & Boumsell, L. (1986) *J. Immunol.* **136**, 1752–1758.
6. Kahn-Perles, B., Wietzerbin, J., Caillol, D. H. & Lemonnier, F. (1985) *J. Immunol.* **134**, 1759–1765.
7. Amiot, M., Dastot, H., Degos, L., Schmid, M. & Boumsell, L. (1987) in *Leucocyte Typing III*, ed. McMichael, A. J. (Oxford Univ. Press, Oxford, U.K.), pp. 80–81.
8. Terhorst, C., van Agthoven, A., LeClair, K., Snow, P., Reinherz, E. & Schlossman, S. (1981) *Cell* **23**, 771–780.
9. Van de Rijn, M., Lerch, P., Knowles, R. W. & Terhorst, C. (1983) *J. Immunol.* **131**, 851–855.
10. Ziegler, A. & Milstein, C. (1979) *Nature (London)* **279**, 243–244.
11. Kefford, R. F., Calabi, F., Fearnley, I. M., Burrone, O. R. & Milstein, C. (1984) *Nature (London)* **308**, 641–642.
12. Bernabeu, C., Van de Rijn, M., Lerch, P. G. & Terhorst, C. (1983) *Nature (London)* **308**, 642–645.
13. Snow, P. M., Van de Rijn, M. & Terhorst, C. (1985) *Eur. J. Immunol.* **15**, 529–532.
14. Ledbetter, J. A., Tsu, T. T. & Clark, E. A. (1985) *J. Immunol.* **134**, 4250–4254.
15. Martin, L. H., Calabi, F., Lefebvre, F. A., Bilsland, C. A. G. & Milstein, C. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 9189–9193.
16. Fithian, E., Kung, G., Goldstein, G., Rubenfeld, M., Fenoglio, C. & Edelson, R. L. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2541–2544.
17. Small, T. N., Knowles, R. W., Keever, C., Kernan, N. A., Collins, N., O'Reilly, R. J., Dupont, B. & Flomenberg, N. (1987) *J. Immunol.* **138**, 2864–2868.
18. Calabi, F. & Milstein, C. (1986) *Nature (London)* **323**, 540–543.
19. Seed, B. & Arruffo, A. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 3365–3369.
20. Martin, L. H., Calabi, F. & Milstein, C. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 9154–9158.
21. Calabi, F., Schroeder, J., Martin, L. H. & Milstein, C. (1987) in *Leucocyte Typing III*, ed. McMichael, A. J. (Oxford Univ. Press, Oxford, U.K.), pp. 72–74.
22. Messing, J. (1983) *Methods Enzymol.* **101**, 20–78.
23. Feinberg, A. P. & Vogelstein, B. (1984) *Anal. Biochem.* **137**, 266–267.
24. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
25. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
26. Auffray, C. & Rougeon, F. (1980) *Eur. J. Biochem.* **107**, 303–314.
27. Kozak, M. (1986) *Cell* **44**, 283–292.
28. Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L. & Wiley, D. C. (1987) *Nature (London)* **329**, 506–512.
29. Kabat, E. A., Wu, T. T., Reid-Miller, M., Perry, H. M. & Gottesman, K. S. (1987) *Sequences of Proteins of Immunological Interest* (U.S. Dept. of Health and Human Services, Washington, D.C.).
30. Hanau, D., Fabre, M., Schmitt, D. A., Garaud, J. C., Pauly, G., Tongio, M. M., Mayer, S. & Cazenave, J. P. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 2901–2905.
31. Amiot, M., Dastot, H., Fabbi, M., Bernard, A. & Boumsell, L. (1987) in *Leucocyte Typing III*, ed. McMichael, A. J. (Oxford Univ. Press, Oxford, U.K.), pp. 81–85.
32. Knowles, R. W. (1987) in *Leucocyte Typing III*, ed. McMichael, A. J. (Oxford Univ. Press, Oxford, U.K.), pp. 86–88.