Immunogenetics 30: 370–377, 1989

Immuno-genetics

The rabbit *CD1* and the evolutionary conservation of the *CD1* gene family

Franco Calabi[1], K. Tertia Belt[1], Chack-Yung Yu[1], Andrew Bradbury[1], William J. Mandy[2], and Cesar Milstein[1]

[1] MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, United Kingdom
[2] Department of Microbiology, University of Texas, Austin, TX 78712, USA

Abstract. A comparison of the genes encoding the CD1 leucocyte differentiation antigens in man and mouse shows important differences which prompted us to analyze the *CD1* genes of the rabbit. We have found that the rabbit genome contains multiple *CD1* loci. Upon cloning and sequencing, one of these loci was found to encode the known rabbit CD1-like antigen (R-Ta) and to be closely related to the human *CD1b* gene, which is absent in the mouse, while a second rabbit gene is closely related to both the human *R3* and the mouse *CD1* genes. The data reinforce the notion of the existence of two classes of *CD1* genes, one of which is conserved in all species, while the other, albeit also evolutionarily old, has been deleted in mice as well as in other rodents.

Introduction

The *CD1* gene family has thus far been studied only in man and in mouse. Five genes in man (Calabi and Milstein 1986, Martin et al. 1986) and two genes in the mouse have been identified (Bradbury et al. 1988). On the basis of sequence analysis, the suggestion has been made that the *CD1* genes fall into two distinct classes (Calabi et al. 1989): the first includes the genes encoding the serologically defined CD1a, -b, and -c antigens; the second consists of one of the remaining human genes (*R3*) as well as of both mouse genes. Thus, the latter class is present in both species, while the former is not, although some functional homologue (i. e., mouse TL) might exist. In this paper, we analyze the *CD1* genes of the rabbit, where a *CD1*-like system has been recently described (Wang et al. 1987).

The nucleotide sequence data reported in this paper have been submitted to the GenBank nucleotide sequence database and have been assigned the accession number M26248 and M26249.
Offprint requests to: C. Milstein.

Materials and methods

Cotton-tail and domestic rabbit genomic libraries were made from partially *Sau* 3AI-digested DNA cloned into the *Bam* HI site of λ EMBL4 and probed with the insert from clone FCB1 as previously described (Martin et al. 1986). Positive clones from the domestic rabbit library were further screened by hybridization overnight at 40 °C to the synthetic oligonucleotide mixture DoRb 10-15 (5′– ATYTGCATXACR-TGRTA –3′, where Y stands for T and C, R for A and G, and X for all four bases), followed by washing three times at room temperature and once at 40 °C in 5 × standard sodium citrate (SSC) + 1% sodium dodecyl sulfate. The following fragments from the λ clones were subject to shotgun cloning (Bankier et al. 1987): 1) from λ CtRb1 (cotton-tail rabbit), the ~2 kbp *Eco* RI-*Bam* HI fragment adjacent to the left arm and the ~1.5 kbp *Bam* HI-*Eco* RI fragment next to it; 2) from λ DoRb3A (domestic rabbit), a 1.7 kbp *Hind* III fragment hybridizing to DoRb 10–15. Shotgun libraries were made in either the *Eco* RV site of TG131 (cotton-tail rabbit) or the *Sma* I site of M13mp8 (domestic rabbit). In the case of the library made from the ~2 kbp *Eco* RI-*Bam* HI fragment of λ CtRb1, only the clones hybridizing to the human *CD1* probe were sequenced. DNA was sequenced by the dideoxy method (Bankier et al. 1987): at least twofold redundancy was achieved on each strand. Percent sequence similarities were determined using the TWOB program (R. Staden, unpublished data). Northern and Southern blots were performed essentially as described (Maniatis et al. 1982) using Hybond-N or Biodyne transfer membranes.

Results

Genomic libraries from both cotton-tail and domestic rabbits were screened with a human *CD1* probe. The cotton-tail rabbit library yielded a clone (λ CtRb1, Fig. 1) con-

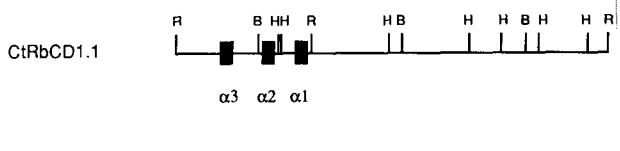


Fig. 1. Restriction map of clone λ CtRb1. R, *Eco* RI; B, *Bam* HI; H, *Hind* III. The approximate position of exons is indicated by *boxes.*

a.

5′UT/Leader

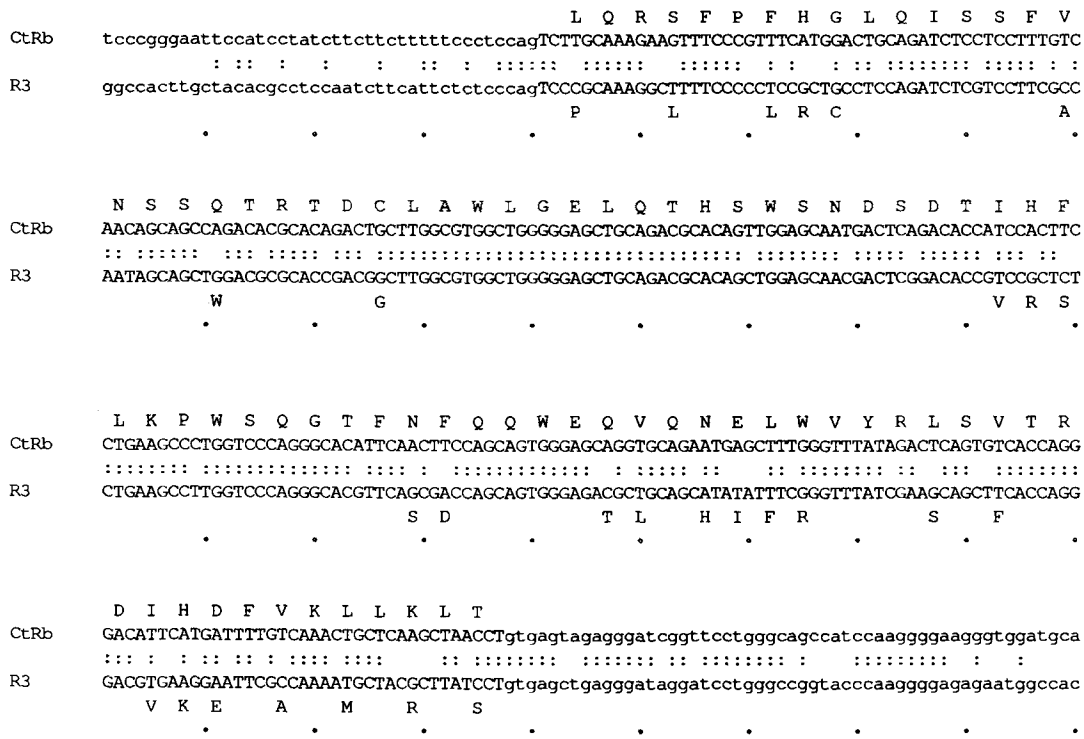
```
DoRb    ggtggttttccaaagtagTGTG-AGGAAGCTGAGGAACCAAATGAAGCCATGATAGTAAGAGATTTAGAAGTTAGAGAT----------
CD1b    ggggcttttcca-aggagGTATGAAGGAAGGTGAGGA-CAGGGAGAGCGGCTGGAAGTCAGGGGGTAAGAGAAACTCTAAAAATCAGGGC

DoRb    GGAGGGGAAATGAGAGGTAGGAGGCAC------AGGGTGGGGGAGCACTTTTTTTTTTGTATACGTAATGTGGATGAGGAAGATACTTGC
CD1b    TTGAGGGAAATAAAAGGTGAGGTAAGAGGCTCAGGGCTGTGGGAGGCACATTTTTCTCTGAAAAGCAGTTTGGATGAGGAAGAGATTTGG

DoRb    CATTCAGAACAGAGAGGAAGTCACTACAGAGGACTGAGAAAAAGGTTTGATAAAACCAGAGATCAAACACCAGCTCTGAGAGTAAGAAGT
CD1b    CAGTTGGAAGAGAGAAGAAGTCACTACAGGGTACTGAGGAAAAGCTTTGCTGAAATTGGAGATCAAATACCAGCTCTGCCAGTAAGAAGT

                        M  L  L  L  P  L  L  L  L  A  G  R  F  P  G  G  D  N  E  D
DoRb    TTTAACTCCTGGTGAAATGCTGCTTCTGCCACTTCTATTACTGGCAGGTCGCTTCCCCGGTGGTGACAATGAGGATGgtaagagtaactc
CD1b    TGCATCTCCCAGTGAAATGCTGCTGCTGCCATTTCAACTGTTAGCTGTTCTCTTTCCTGGTGGTAACAGTGAACATGgtaagagtcactc
                                        F  Q           V  L              N  S     H
```

α1

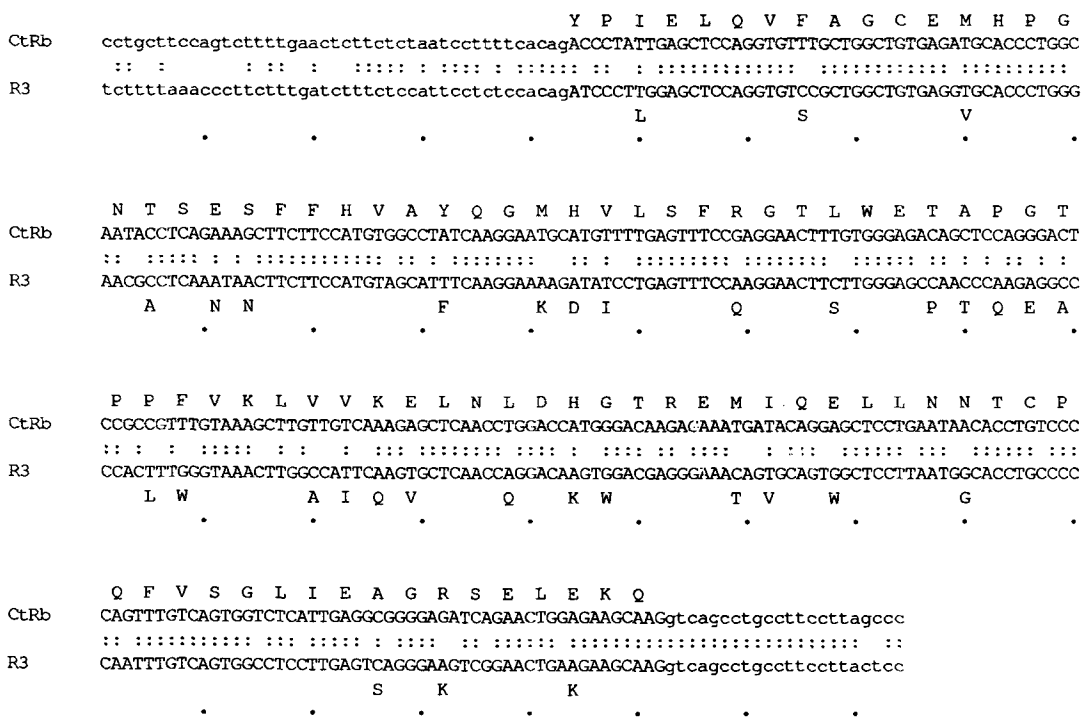
```
                           A  L  Q  G  P  T  S  Y  H  V  M  Q  I  S  S  F  T  N  S  T  W  T  E  N
DoRb    ggccttcctctctttgcagCCCTCCAGGGGCCGACGTCTTACCATGTCATGCAGATCTCTTCCTTTACAAACAGCACCTGGACAGAAAAT
CD1b    tgttttccactcttcacagCCTTCCAGGGGCCGACCTCCTTTCATGTTATCCAGACCTCGTCCTTTACCAATAGTACCTGGGCACAAACT
                             F                 F        I        T                          A  Q  T

         R  G  S  G  W  L  E  D  L  Q  I  H  R  W  D  S  E  T  G  T  A  I  F  L  K  P  W  S  K  G
DoRb    CGAGGCTCAGGCTGGCTGGAAGATTTGCAGATTCACAGGTGGGATAGTGAAACGGGCACTGCCATCTTCCTGAAGCCCTGGTCTAAGGGA
CD1b    CAAGGCTCAGGCTGGTTGGATGATTTGCAGATTCATGGCTGGGATAGCGACTCAGGCACTGCCATATTCCTGAAGCCTTGGTCTAAAGGT
         Q                 D                 G           D  S

         N  L  S  D  E  E  I  T  E  L  V  E  L  F  R  V  Y  F  F  G  L  V  R  E  L  R  D  H  V  T
DoRb    AACTTAAGCGATGAGGAAATTACTGAGCTGGTGGAGCTGTTCCGGGTCTACTTCTTTGGACTCGTTCGAGAACTGCGAGACCACGTCACT
CD1b    AACTTTAGTGATAAGGAGGTTGCTGAGTTAGAGGAGATATTCCGAGTCTACATCTTTGGATTCGCTCGAGAAGTACAAGACTTTGCCGGT
                  K     V  A        E     I              I        F  A        V  Q     F  A  G

         E  F  Q  M  K
DoRb    GAATTCCAGATGAAGTgtgagtccagccttctcccttgggga
CD1b    GATTTCCAGATGAAATgtgagtctagtccaccgaactgggaa
         D
```

b.

α1


L Q R S F P F H G L Q I S S F V
CtRb tcccgggaattccatcctatcttcttcttttccctccagTCTTGCAAAGAAGTTTCCCGTTTCATGGACTGCAGATCTCCTCCTTTGTC
R3 ggccacttgctacacgcctccaatcttcattctctcccagTCCCGCAAAGGCTTTTCCCCCTCCGCTGCCTCCAGATCTCGTCCTTCGCC
P L L R C A
N S S Q T R T D C L A W L G E L Q T H S W S N D S D T I H F
CtRb AACAGCAGCCAGACACGCACAGACTGCTTGGCGTGGCTGGGGGAGCTGCAGACGCACAGTTGGAGCAATGACTCAGACACCATCCACTTC
R3 AATAGCAGCTGGACGCGCACCGACGGCTTGGCGTGGCTGGGGGAGCTGCAGACGCACAGCTGGAGCAACGACTCGGACACCGTCCGCTCT
W G V R S
L K P W S Q G T F N F Q Q W E Q V Q N E L W V Y R L S V T R
CtRb CTGAAGCCCTGGTCCCAGGGCACATTCAACTTCCAGCAGTGGGAGCAGGTGCAGAATGAGCTTTGGGTTTATAGACTCAGTGTCACCAGG
R3 CTGAAGCCTTGGTCCCAGGGCACGTTCAGCGACCAGCAGTGGGAGACGCTGCAGCATATATTTCGGGTTTATCGAAGCAGCTTCACCAGG
S D T L H I F R S F
D I H D F V K L L K L T
CtRb GACATTCATGATTTTGTCAAACTGCTCAAGCTAACCTgtgagtagagggatcggttcctgggcagccatccaaggggaagggtggatgca
R3 GACGTGAAGGAATTCGCCAAAATGCTACGCTTATCCTgtgagctgagggataggatcctgggccggtacccaaggggagagaatggccac
V K E A M R S

α2


Y P I E L Q V F A G C E M H P G
CtRb cctgcttccagtcttttgaactcttctctaatccttttcacagACCCTATTGAGCTCCAGGTGTTTGCTGGCTGTGAGATGCACCCTGGC
R3 tcttttaaaccctttctttgatctttctccattcctctccacagATCCCTTGGAGCTCCAGGTGTCCGCTGGCTGTGAGGTGCACCCTGGG
L S V
N T S E S F F H V A Y Q G M H V L S F R G T L W E T A P G T
CtRb AATACCTCAGAAAGCTTCTTCCATGTGGCCTATCAAGGAATGCATGTTTTGAGTTTCCGAGGAACTTTGTGGGAGACAGCTCCAGGGACT
R3 AACGCCTCAAATAACTTCTTCCATGTAGCATTTCAAGGAAAAGATATCCTGAGTTTCCAAGGAACTTCTTGGGAGCCAACCCAAGAGGCC
A N N F K D I Q S P T Q E A
P P F V K L V V K E L N L D H G T R E M I Q E L L N N T C P
CtRb CCGCCGTTTGTAAAGCTTGTTGTCAAAGAGCTCAACCTGGACCATGGGACAAGACAAATGATACAGGAGCTCCTGAATAACACCTGTCCC
R3 CCACTTTGGGTAAACTTGGCCATTCAAGTGCTCAACCAGGACAAGTGGACGAGGGAAACAGTGCAGTGGCTCCTTAATGGCACCTGCCCC
L W A I Q V Q K W T V W G
Q F V S G L I E A G R S E L E K Q
CtRb CAGTTTGTCAGTGGTCTCATTGAGGCGGGGAGATCAGAACTGGAGAAGCAAGgtcagcctgccttccttagccc
R3 CAATTTGTCAGTGGCCTCCTTGAGTCAGGGAAGTCGGAACTGAAGAAGCAAGgtcagcctgccttccttactcc
S K K

α3

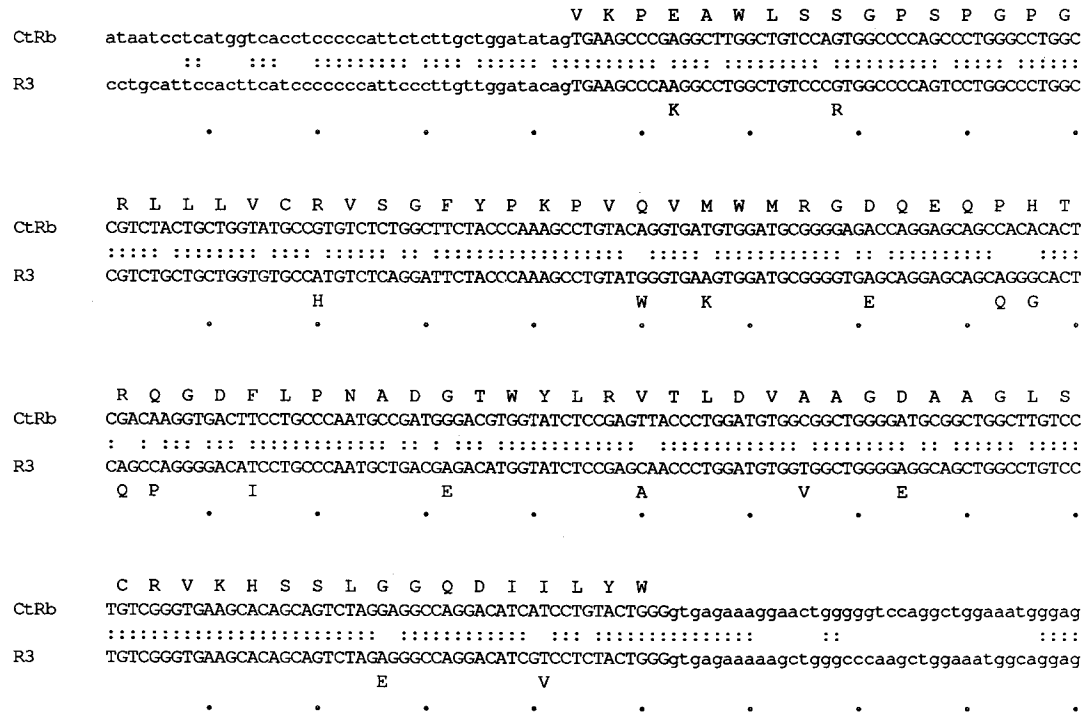


Fig. 2a, b. Nucleotide and predicted amino acid sequences of rabbit *CD1* genes. The domestic rabbit gene is aligned to human *CD1b* (**a**) and the cotton-tail rabbit gene to human *R3* (**b**). *Dashes* have been introduced to maximize alignments. Exons are in *capital*, introns in *small letters*. DNA identities are marked by a *colon*, while in the human sequences only the amino acid residues which are different from the rabbit are indicated. *Boxes* highlight the conserved heptamers GAAGTCA and GGGAAAT (Calabi et al. 1989). *Dots* mark every tenth residue.

taining a single *Eco* RI fragment which hybridized to the human probe. This fragment was subcloned and partially sequenced. The domestic rabbit library yielded several clones, which preliminary restriction mapping analysis identified as belonging to independent loci. Some of these clones also hybridized to a mouse α1 probe (data not shown). The domestic rabbit clones were then screened with a synthetic oligonucleotide mixture complementary to all sequences that could possibly encode residues 10–15 (exept for the third position of the last codon) of the known rabbit CD1 polypeptide sequence (Wang et al. 1988). Only one clone (λ DoRb3A) was found to be positive, and a relevant restriction fragment was subcloned and sequenced.

Figure 2 shows the sequences of the *CD1* exons identified in both cotton-tail and domestic rabbits based on homology to the known genes and on the location of potential splice sites. Only the 5′untranslated/leader and α1 exons have been sequenced in the domestic rabbit and only the α1, α2, and α3 exons from the cotton-tail rabbit. The sequence of the domestic rabbit gene agrees with the published sequence of rabbit *CD1* (R-Ta) at 17 out of 18 positions (Wang et al. 1988). The single discrepancy (at the amino-terminal residue) might be due to either isotypic or allelic differences. However, the N-terminal amino acid is often unreliable, and on re-examination of the original sequence data a minor peak corresponding to glutamic acid was detected at position 1. Therefore, it is most likely that the domestic rabbit gene encodes the known rabbit CD1 antigen and that glutamic acid is its amino terminal residue.

Sequence comparison (Fig. 2 and Table 1) clearly shows the cotton-tail rabbit *CD1* gene to be most related to the human *R3* and mouse *CD1* genes, while the domestic rabbit *CD1* gene is most similar to the human *CD1b* gene. In the latter case, similarly to what has been previously noted in the comparison among the human *CD1a, -b, -c*, and *R2* genes (Martin et al. 1987, Calabi et al. 1989), the homology extends to the whole of the 5′untranslated/leader exon, up to and beyond a proposed 5′ splice acceptor site; in particular, two blocks of sequence, which are found in all four human genes, including the heptamers GAAGTCA and GGGAAAT and located ~90 and ~180 basepairs upstream of the proposed translational start, are also conserved in sequence and position in the domestic rabbit gene (Fig. 3). From an alignment of all available *CD1* sequences (Fig. 4), amino acid residues emerge which appear to identify either the

Table 1. Percent sequence similarities among *CD1* genes and polypeptides.

Leader

	b		c		DoRb1		R2		R3		M1		M2		CtRb1	
a	63	**39**	70	**56**	67	**50**	44	**28**	50	**44**	44	**39**	43	**39**	–	–
b			74	**61**	74	**67**	59	**44**	50	**39**	41	**28**	39	**28**	–	–
c					72	**67**	56	**50**	59	**44**	46	**33**	48	**33**	–	–
DoRb1							59	**50**	48	**28**	46	**33**	44	**33**	–	–
R2									54	**33**	41	**22**	43	**22**	–	–
R3											69	**61**	70	**61**	–	–
M1													93	**94**	–	–
M2															–	–

α1

	b		c		DoRb1		R2		R3		M1		M2		CtRb1	
a	64	**51**	70	**58**	64	**53**	63	**50**	54	**36**	51	**36**	49	**34**	50	**35**
b			70	**62**	80	**73**	64	**53**	59	**46**	58	**42**	56	**39**	54	**39**
c					64	**57**	69	**54**	57	**42**	56	**39**	54	**35**	54	**39**
DoRb							63	**49**	58	**42**	58	**39**	54	**35**	54	**36**
R2									64	**46**	63	**49**	61	**46**	57	**39**
R3											78	**64**	77	**64**	77	**70**
M1													95	**91**	72	**55**
M2															70	**54**

α2

	b		c		DoRb1		R2		R3		M1		M2		CtRb1	
a	60	**38**	64	**39**	–	–	62	**41**	57	**34**	54	**34**	52	**32**	59	**32**
b			65	**50**	–	–	59	**43**	54	**34**	50	**32**	48	**29**	52	**27**
c					–	–	56	**37**	56	**38**	51	**35**	50	**34**	55	**37**
DoRb1							–	–	–	–	–	–	–	–	–	–
R2									59	**39**	53	**36**	51	**33**	59	**34**
R3											72	**55**	71	**55**	72	**54**
M1													95	**91**	65	**53**
M2															68	**51**

α3

	b		c		DoRb1		R2		R3		M1		M2		CtRb1	
a	88	**85**	76	**72**	–	–	87	**82**	86	**78**	78	**72**	78	**72**	77	**72**
b			76	**70**	–	–	88	**82**	87	**78**	78	**71**	78	**71**	79	**72**
c					–	–	75	**67**	76	**66**	75	**62**	75	**62**	73	**63**
DoRb1							–	–	–	–	–	–	–	–	–	–
R2									89	**93**	81	**73**	80	**73**	80	**77**
R3											83	**75**	83	**75**	85	**80**
M1													98	**96**	79	**72**
M2															80	**72**

Percent sequence similarities amongst *CD1* genes and polypeptides were calculated on the block alignments shown in Figure 4 rather than on optimal pairwise comparisons, which explains the minor differences with Calabi and co-workers (1988). Plain type refers to DNA, bold type to protein. Boxes highlight clusters within which DNA identity is >60% and amino acid identity is >38% among any two members.

whole *CD1* family or one of the two classes. The latter are found only in the α1 and α2 domains and include the majority of potential N-glycosylation sites as well a presumably unpaired cysteine in the α2 domain.

In order to determine the complexity of the rabbit *CD1* gene family, Southern blots of domestic rabbit DNA were probed with a human α3 *CD1* probe. The results (Fig. 5) suggest the existence of up to eight genes. Only one of these, however, was identified as a member of the *R3* class, as shown by hybridization at high stringency with an α1 probe from the *R3*-like cotton-tail rabbit gene described above.

Northern blotting analysis (Fig. 6) shows rabbit *CD1* to be expressed in the thymus, but not in bone marrow, spleen, liver, brain, or testis. These data are consistent with the known tissue distribution of the rabbit CD1 antigen (R-Ta; Wang et al. 1987). The size of the main transcripts is similar to that described in both humans and mice (Calabi and Milstein 1986, Bradbury et al. 1988).

```
                -183                    -94
CD1a (man)        AGGGAAATGAGA.......AAAGAAGTCAGAATA
                -188                    -90
CD1b (man)        AGGGAAATAAAA.......GAAGAAGTCACTACA
                -180                    -90
CD1c (man)        GGGGAAATGAGA.......AAGGAAGTCAGAATA
                -201                    -102
R2 (man)          AGGGAAATGAAA.......AGGGAAGTCAGACGA
                -190                    -93
R-Ta (rabbit)     GGGGAAATGAGA.......GAGGAAGTCACTACA
-------------------------------------------------------
Consensus         A      G    A G      GA
                  GGGGAAATGA A A.....G A AGAAGTCA CT A-A
```

Fig. 3. Sequence conservation in *CD1* 5′ regions. *Numbers* refer to the position with respect to the proposed translational start. The consensus consists of bases conserved in at least four out of the five sequences.

Discussion

Genes which cross-hybridize with human *CD1* are present in multiple copies in the rabbit genome. One of them has been identified as encoding the previously described R-Ta antigen, recognized by mAb 5E2 (Wang et al. 1987). This gene has been partially sequenced, and comparison with previously described *CD1* genes shows highest similarity to human *CD1b*. A second gene from the cotton-tail rabbit has also been sequenced and shown to be closely related to human *R3* (a gene for which a polypeptide product has not yet been identified) and, to a slightly lesser extent,

Leader

```
                -10
CD1a        MLFLLLPLL-AVLPG-DGN
CD1b        MLLLPFQLLAVLFPG-GNS
CD1c        MLFLQFLLLALLLPG-GDN
DoRbCD1.1   MLLLPLLLLAGRFPG-GDN

R2          MLLL-FLLFEGL■■P-GEN

R3          MGCLLFLLLWALLQAWG-S
MCD1.1      MRYLPWLLLWAFLQVWGQQ
MCD1.2      MRYLP■LLLWAFLQVWGQQ
```

α1 domain

```
                    10        20        30        40        50        60        70        80        90
CD1a        ADGLKEPLSFHVTWIASFYFHSWKQNLVSGWLSDLQTHTWDSISSTIVFLPWSRGEFSNEEWKELETLFRIRTIRSFEGIRRYAHELQFE--
CD1b        EHAFQGPTSFHVIQTSSFTFSTWAQTQGSGWLDDLQIHGWDSDSGTAIFLKPWSKGEFSDKEVAELEEIFRVYIFGFAREVQDFAGDFQMK--
CD1c        ADASQEHVSFHVIQIFSFVPQSWARGQGSGWLDELQTHGWDSESGTIIFLHNWSKGEFSNEELSDLELLFRFYLFGLTREIQDHASQDYSK--
DoRbCD1.1   EDALQGPTSYHVMQISSFTFSTWTENRGSGWLEDLQIHRWDSETGTAIFLKPWSKGELSDEEITELVELFRVYFFGLVRELRDHVTEFQMK--

R2          TAAAEEQLSFRMLQTSSFAHSWAHSEGSGWLGDLQTHGWDTVLGTIRFLKPWSHGEFSKQELKNLQSLFQLYFHSFIQIVQASAGQFQLE--

R3          AEVPQRLFPLRLLQISSFASSWTRTDGLAWLGELQTHSWSIDSDTVRSLKPWSQGTFSDQQWETLQHIFRVYRSSFTRDVKEFAKMLRLS--
MCD1.1      SEAQQKIYTFRILQMSSFASRSWSRTDSVVWLGDLQTHRWSIDSATISFTKPWSQGKLSNQQWEKLQHMFQVYRVSFTRDIQELVKMMSPKED
MCD1.2      SEVQQKIYTFRILQTSSFASISWSRTDSLILLGDLQTHRWSIDSAIISFTKPWSQGKLSNQQWEKLQHMFQVYRVSFTRDIQELVKMMSPKED
CtRbCD1.1      LQRSFPFHGLQISSFVPSSQTRTDILAWLGELQTHSWSIDSDTIHFLKPWSQGTFNFQQWEQVQNELWVYRLSVTRDIHDFVKLLKLT--
```

α2 domain

```
                100       110       120       130       140       150       160       170       180
CD1a        YPFEIQVTGLELHSGKVSGSFLQLAYQGSDFVSFQGNSWLPYPVAGNMAKHFCKVLN-QNQHENDITHNLLSDTLPRFILGLLDAGKAHLQRQ
CD1b        YPFEIQGIAGELHSGGAIVSFLRGALGGLDFLSVKEASLVPSPEGGSRAQKFLALII-QYQGIMETVRILLYETLPRYLLGVLNAGKADLQRQ
CD1c        YPFEVQVKAGELHSGKSPEGFFQVAFNGLDLLSFQGTTWVPSPGLGSLAQSVCHLLNHQYEGVTETVYNLIRSTLPRFLLGLLDAGKMYVHRQ

R2          YPFEIQILAGLRM---NAPQIFLNMAYQGSDFLSFQGISWEPSPGAGIRAQNICKVLN-RYLDIKEILQSLLGHTLPRFLAGLMEAGESELKRK

R3          YPLELQVSAGLEVHPGSASNNFFHVAFQGKDILSFQGTSWEPTQEAPLWVNLAIQVLN-QDKWTRETVQWLLKGTLPQFVSGLLESGKSELKKQ
MCD1.1      YPIEIQLSAGLEMYPGSASESFLHVAFQGKYVVRFWGTSWQTVPGAPSWLDLPIKVLN-ADQGTSATVQMLLKDTLPLFVRGLLEAGKSDLEKQ
MCD1.2      YPIEIQLSTGLEMYPGSASESFFHVAFQGKYAVRFRGTSWQRVLGAPSWLDLPIKVLN-ADQGTSATVQTLLKDTWPQFARGLLEAGKSDLEKQ
CtRbCD1.1   YPIELQVFAGLEMHPGSTSESFFHVAYQGMHVLSFRGTLWETAPGTPPFVKLVVKELN-LDHGTREMIQELLKNTLPQFVSGLIEAGRSELEKQ
```

α3 domain

```
                190       200       210       220       230       240       250       260       270
CD1a        VKPEAWLSHGPSPGPGHLQLVLHVSGFYPKPVWVMWMRGEQEQQGTQRGDILPSADGTWYLRATLEVAAGEAADLSCRVKHSSLEGQDIVLYW
CD1b        VKPEAWLSSGPSPGPGRLQLVLHVSGFYPKPVWVMWMRGEQEQQGTQLGDILPNANWTWYLRATLDVADGEAAGLSCRVKHSSLEGQDILLYW
CD1c        VRPEAWLSSRPSLGSGQLLLVLHASGFYPKPVWVTWMRNEQEQLGTKHGDILPNADGTWYLQVILEVASEEPAGLSCRVRHSSLGGQDIILYW

R2          VKPEAWLSCGPSPGPGRLQLVLHVSGFYPKPVWVMWMRGEQEQRGTQRGDVLPNADETWYLRATLDVAAGEAAGLSCRVKHSSLGGHDLIIHW

R3          VKPKAWLSRGPSPGPGRLLLVLHVSGFYPKPVWVKWMRGEQEQQGTQPGDILPNADETWYLRATLDVVAGEAAGLSCRVKHSSLEGQDIVLYW
MCD1.1      EKPVAWLSSVPSSAHGHRQLVLHVSGFYPKPVWVMWMRGDQEQQGTHRGDFLPNADETWYLQATLDVEAGEEAGLACRVKHSSLGGQDIILYW
MCD1.2      EKPVAWLSSVPSSAHGHLQLVLHVSGFYPKPVWVMWMRGDQEQQGTHRGDFLPNADETWYLQATLDVEAGEEAGLACRVKHSSLGGQDIILYW
CtRbCD1.1   VKPEAWLSSGPSPGPGRLLLVLHVSGFYPKPVQVMWMRGDQEQPHTRQGDFLPNADGTWYLRVTLDVAAGDAAGLSCRVKHSSLGGQDIILYW
```

Fig. 4. Alignment of all available *CD1* sequences. *Numbering* is based on *CD1a*. Potential N-glycosylation sites are *boxed*, and cysteine residues are *highlighted*. Constant residues (present in at least eight of the nine sequences) and class-specific residues (present in all and only the sequences of a given class as well as possibly in the human *R2* gene, which is regarded as lying in an intermediate position) are indicated by *different shading*.

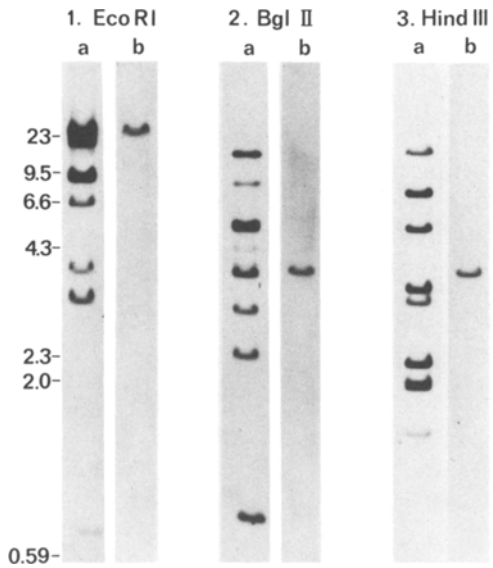


Fig. 5. Southern blotting analysis of genomic DNA from domestic rabbit. The restriction endonucleases used are indicated. The blot was sequentially hybridized (a) to a human *CD1* probe, which cross-hybridizes to all *CD1* genes [a mixture of inserts from clone FCB1 (Calabi and Milstein 1986) and from an M13 shotgun clone spanning the $\alpha 3$ domain of *CD1c*] and (b), after stripping the first probe, to the 0.8 kb *Eco* RI-*Hind* III fragment from λ CtRb1 which contains R-3-class $\alpha 1$ exon. Washing was performed twice in 6 × SSC at room temperature, twice in 6 × SSC at 65 °C, and twice in 6 × SSC at room temperature. The positions of reference size markers (in kbp) are indicated.

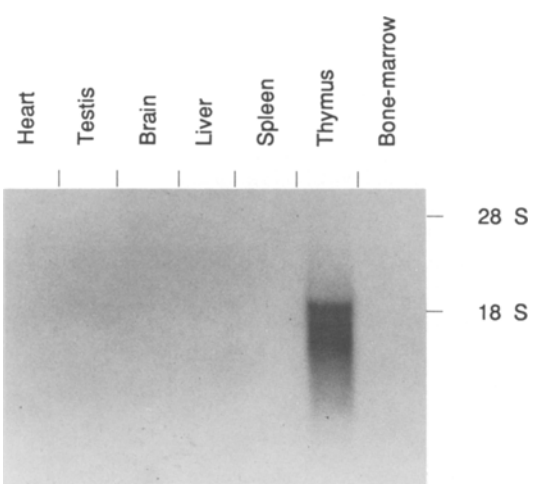


Fig. 6. Northern blotting analysis of total RNA from the indicated tissues from domestic rabbit. All samples contained approximately equal amounts of RNA (~20 μg) as judged by ethidium bromide staining. The probe was the insert of an M13 shotgun clone spanning the cottontail rabbit $\alpha 3$ exon (clone CtRbSG103). Identical results were obtained with the 0.8 kbp *Eco* RI-*Hind* III fragment λ CtRb1 which contains the $\alpha 1$ exon. Washing was performed three times at 65 °C in 2 × SSC. The position of reference rRNA markers is indicated.

to both mouse genes. Moreover, hybridization data with a class-specific probe clearly show the existence in the rabbit of both *CD1* classes that we have previously identified in humans (Calabi et al. 1989). Only *R3*-class genes, however, have been found in the mouse. While we have previously suggested that the origin of the *R3*-class genes predates the radiation of mammalian species (~100 million years ago), the data presented in this paper imply the same conclusion for the *CD1a, -b,* and *-c* class, as the latter genes are found in both man and rabbit, albeit not in the mouse. The similarity between the rabbit and the respective human sequences is considerable and shows that these genes are not evolving at a particularly fast rate. Therefore, the absence of the *CD1a, -b, -c* cluster in the mouse must have been due to a deletion taking place after the separation between the mouse and the rabbit/human lineages.

Although it remains unknown what, if any, the function of *CD1* is, the conservation of amino acid sequences does suggest a functional significance. Likewise, the differential role that the two *CD1* classes may play remains to be established. A functional difference is, however, suggested by the finding of class-specific potential glycosylation sites and intermolecular disulfides.

The R-Ta antigen has only been found expressed in the thymus (Wang et al. 1987). Northern blotting analysis shows no rabbit *CD1* transcript outside the thymus. While this pattern of expression closely resembles that known for human *CD1* (Bernard et al. 1984), mouse *CD1* has been shown to be transcribed in several extrathymic tissues (Bradbury et al. 1988).

Thus, man and rabbit *CD1* are closer to each other than either is to the mouse not only in sequence, in keeping with accepted phylogenetic relationships, but also in overall genetic organization and in pattern of expression. For functional studies on *CD1*, therefore, the rabbit provides an experimental animal model which is most relevant to the human system.

Acknowledgments. We thank Katherine Knight (University of Illinois Medical School, Chicago) for kindly providing the domestic rabbit genomic library, Rodger Staden for computer programs, and John J. Jarvis for expert technical assistence. F. C. was supported by a special fellowship of the Leukemia Society of America, T. B. by a Beit Memorial Research Fellowship, C.Y.Y. by a fellowship from the Croucher Foundation, and A.B. by a MRC Research Fellowship.

References

Bankier, A. T., Weston, K. M., and Barrell, B. G.: Random cloning and sequencing by the M13/dideoxynucleotide chain termination method. *Methods Enzymol 155:* 51-93, 1987

Bernard, A., Boumsell, L., Dausset, J., Milstein, C., and Schlossman, S. F. (eds.): *Leucocyte Typing*, Springer-Verlag, Berlin, 1984

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

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

Maniatis, T., Fritsch, E. F., and Sambrook, J.: *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, New York, 1982

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

Martin, L. H., Calabi, F., Lefebvre, F.-A., Bilsland, C. A. G., and Milstein, C.: Structure and expression of the human thymocyte antigens CD1a, CD1b and CD1c. *Proc Natl Acad Sci USA 84:* 9189–9193, 1987

Wang, C.-R., Chen, G.-H., and Mandy, W. J.: Identification of a rabbit class I like thymocyte specific antigen. *J Immunol 18:* 3352–3359, 1987

Wang, C.-R., Chen, G.-H., Newkirk, M., Capra, J. D., and Mandy, W. J.: Biochemical properties of a novel rabbit thymocyte specific class-I like antigen. *Mol Immunol 25:* 945–952, 1988

Received June 22, 1989.