M. Han
L.I. Hannick
M. DiBrino
M.A. Robinson

Polymorphism of human CD1 genes

Key words:
CD1; cell surface molecules; MHC class I; molecular biology

Acknowledgments:
The authors wish to thank Ms. Nancy Cogan for secretarial assistance.

Abstract: Human CD1 genes have been reported to be invariant or to show limited polymorphism. Recently, certain functions of CD1 antigens have been described to include the presentation lipid and glycolipid antigens. These observations prompted a thorough survey of the genetic polymorphism in the five human CD1 genes (CD1a–CD1e). Using polymerase chain reaction-single stranded conformational polymorphism (PCR-SSCP) combined with sequence analyses, exons 2 and 3 from CD1a–CD1e were characterized from a total of 110 unrelated healthy donors. Results showed that all five genes (CD1a–CD1e) are polymorphic in exon 2. Substitutions in CD1b and CD1c are silent, whereas, substitutions in CD1a, CD1d and CD1e result in amino acid replacements in the deduced protein products. CD1a and CD1e polymorphisms are prevalent in the population. The substitutions in CD1a have characteristics that may influence interactions with β_2-microglobulin (β_2-m) or accessory molecules. The substitution in CD1e is located in the region predicted to interact with ligands and may differentially impact the ability of CD1e alleles to bind antigen.

Authors' affiliation:

M. Han,
L.I. Hannick,
M. DiBrino,
M.A. Robinson

Laboratory of Immunogenetics, National Institute of Allergy and Infectious Diseases, Twinbrook II Facility, National Institutes of Health, Rockville, Maryland, USA

Correspondence to:
Mary Ann Robinson, Ph.D.
Twinbrook II Facility
12441 Parklawn Drive
Rockville, MD 20852
USA
Tel: +1 301 496 9888
Fax:+1 301 402 0259
e-mail: mr23k@nih.gov

CD1 molecules are cell surface glycoproteins that are structurally and functionally similar to major histocompatibility complex (MHC) class I molecules (1, 2). Both CD1 and class I genes have undergone duplication events to generate a series of linked genes (3). There are five CD1 genes in humans (CD1a, b, c, d, and e) and two in mice (CD1d1.1 and CD1d1.2) (4). CD1 and class I genes have similar genomic organizations and encode molecules composed of a heavy chain molecule with three extracellular domains, α1, α2, and α3 (1). Crystallographic structures of class I molecules and of murine CD1d revealed that the α1 and α2 domains form a ligand binding groove and the α3 domains associate noncovalently with β_2-microglobulin (β_2-m) (5).

A primary structural difference between CD1 and class I molecules is in the characteristics of the ligand binding groove. Class I molecules bind a variety of peptide antigens and sequence differences among class I molecules can determine the features of pep-

Received 14 April,
accepted for publication 17 May 1999

Tissue Antigens . ISSN 0001-2815

Tissue Antigens 1999: **54**: 122–127
Printed in Denmark . All rights reserved

tides that can be presented (6). In contrast to class I molecules, the α1 and α2 domains of CD1 are hydrophobic and the shape of the ligand binding groove is narrower and deeper (5). Information concerning CD1 ligands has been obtained from the definition of the specificity of double negative and CD8+ T lymphocytes that recognize lipid or glycolipid antigens presented by CD1 (7–10).

In the present study, polymorphism in CD1 genes was examined in a large panel of healthy individuals. Alleles of CD1a–CD1e were identified by using single stranded conformational polymorphism (SSCP) (11) and sequence analysis. The present study shows that all five CD1 genes are polymorphic in the α1 domain. Both allelic forms of CD1a and CD1e were prevalent in the population and the substitutions resulted in amino acid replacements that may influence the function of CD1a and CD1e alleles.

Material and methods

Genomic DNA

Genomic DNA samples were isolated from B lymphoblastoid cell lines or from PBL as previously described (12). Cell donors included 110 unrelated individuals representing diverse ethnic backgrounds. Family members of individuals with informative markers were studied; segregation analyses were performed in 5 families.

Single stranded conformational polymorphism (SSCP) DNA samples were polymerase chain reaction (PCR) amplified using combinations of exon 2 forward and reverse primers shown by underlining in Fig. 1. Primers for exon 3 were CD1a.ex3F1 CAC TCT GGA AAG GTC TCA GG; CD1b.ex3F1 GTG CCA TAG TAA GCT TCC TG; CD1c.ex3F1 CTG GAA AGA GCC CAG AAG GC; CD1d.ex3F1 CTG GGA ACG CCT CAA ATA AC; CD1e.ex3F1 CTG TAG AAT GAA TGC CCC AC; CD1a.ex3R1 GCT GGA GAT GTG CCT TTC CT; CD1b.ex3R1 GCA GAT CTG CTT TTC CTG CA; CD1c.ex3R1 GTG TAC ATA CAT CTT CCC TG; CD1d.ex3R1 TCA GTT CCG ACT TCC CTG AC; CD1e.ex3R1 AGT TCT GAC TCC CCT GCT TC. The PCR mixture contained 1 X PCR buffer, 0.2 mM each of four deoxynucleotides, 1.5 mM MgCl2, 1 mM each CD1 forward and reverse primers, 0.5 U Taq polymerize, 1 mCi ^{32}P-dCTP, and 0.2 mg genomic DNA in a total volume of 20 μl. The PCR reaction was performed at 95°C, 55°C, and 72°C for 30 s., 30 s. and 90 s. respectively, and finally incubated for 7 min at 72°C. Prior to electrophoresis, DNA samples were diluted 1:10 with 0.1% SDS, 10 mM EDTA at pH 8 and then diluted again 1:2 with sequencing stop solution (Amersham, Arlington Heights, IL, USA). The sample was denatured by incubation at 95°C for 5 min, then immediately placed on ice before application to the non-denaturing SSCP gel. The SSCP gel was composed of 0.6 X TBE, 10% glycerol, 3% long ranger gel solution, 0.05% Ammonium Persulfate and 0.05% TEMED. Electrophoresis was performed at 30 W for 4 h. Gels were dried on filter papers and exposed to X-ray film at −80°C for 24–48 h with an intensifying screen.

Sequence analysis

Based upon SSCP patterns, DNA samples were selected for sequence analysis. PCR reactions were performed as above but without ^{32}P-dCTP. PCR products were ligated in the pCRII vector and

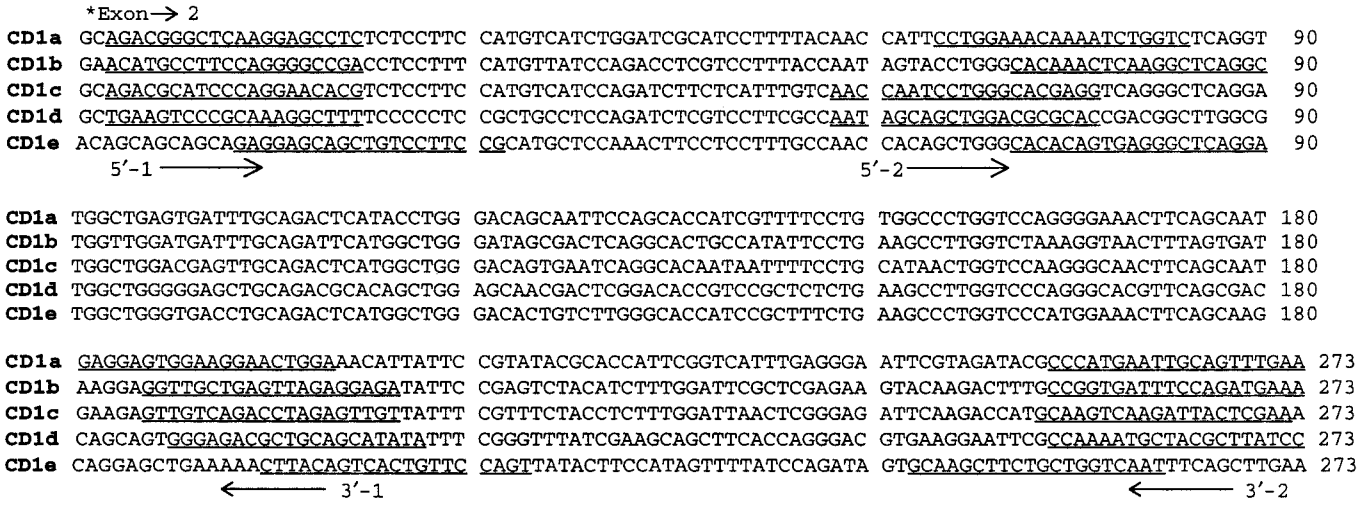

Fig. 1. **Alignment of CD1 exon 2 nucleotide sequences.** *Underlines* indicate the positions that correspond to PCR primers. Accession numbers for each CD1 subfamily gene are from GenBank: CD1a: M28825; CD1b: M28826; CD1c: M28827; CD1d: J04142, and CD1e: X14975.

transformed into competent *Escherichia coli* INVaF′ (TA Cloning Kit; Invitrogen, Carlsbad, CA, USA). Recombinant plasmids were expanded and DNA samples were purified with QIAwell 8 plus plasmid kit (Qiagen, Santa Clarita, CA, USA). The eluted DNA was subjected to a cycle sequencing reaction using the DNA sequencing kit (Perkin-Elmer, Foster City CA, USA). PCR products were purified by AGTC Gel Cartridges (Edge Biosystems, Gaithersburg, MD, USA) and sequences were determined using an ABI 377 Sequencer. A total of 126 clones derived from 18 different individuals were sequenced; at least 15 clones were sequenced for each CD1 gene.

Molecular modeling

All of the modeling was done by L.I.H. using the program RasMol (13) for Fig. 3A–B. Three-dimensional coordinates are based on the crystal structure of murine CD1 (5).

Results

Polymorphism in exon 2 and exon 3 encoding the α1 and α2 domains of human CD1 was examined using SSCP (11), a technique that can reveal sequence variations of even a single nucleotide. Sequences for exon 2 of the 5 human CD1 genes, CD1-a, -b, -c, -d, and -e, were retrieved from GenBank and aligned by using the MAP multiple sequence alignment program (Fig. 1). Two sets of primers overlapping exon 2 (5′-1 with 3′-1 and 5′-2 with 3′-2) sequences and one set of primers for exon 3 were used to examine DNA samples from 110 healthy unrelated individuals. All 5 CD1 genes showed migration polymorphisms for exon 2, which correspond to the α1 domain of CD1. No variations in migration patterns were observed for exon 3.

Two allelic forms were revealed by SSCP analysis for each of the CD1 genes and DNA sequence analyses were performed on in-

```
CD1a                                                13
             L   K   E   P   L   S   F   H   V   I   W   I   A   S   F   Y   N   H   S   W
Allele 1    CTC AAG GAG CCT CTC TCC TTC CAT GTC ATC TGG ATC GCA TCC TTT TAC AAC CAT TCC TGG   69
Allele 2    --- --- --- --- --- --- --- --- --- -C- --- --- --- --- --- --- --- --- --- --->
             -   -   -   -   -   -   -   -   -   T   -   -   -   -   -   -   -   -   -   -

             K   Q   N   L   V   S   G   W   L   S   D   L   Q   T   H   T   W   D   S   N
Allele 1    AAA CAA AAT CTG GTC TCA GGT TGG CTG AGT GAT TTG CAG ACT CAT ACC TGG GAC AGC AAT  129
Allele 2    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --->
             -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -
                                     51
             S   S   T   I   V   F   L   W   P   W   S   R   G   N   F   S   N   E   E   W
Allele 1    TCC AGC ACC ATC GTT TTC CTG TGG CCC TGG TCC AGG GGA AAC TTC AGC AAT GAG GAG TGG  189
Allele 2    --- --- --- --- --- --- --- --C --- --- --- --- --- --- --- --- --- --- --- --->
             -   -   -   -   -   -   -   C   -   -   -   -   -   -   -   -   -   -   -   -

CD1b                                                                        58
             D   S   G   T   A   I   F   L   K   P   W   S   K   G   N   F   S   D   K   E
Allele 1    GAC TCA GGC ACT GCC ATA TTC CTG AAG CCT TGG TCT AAA GGT AAC TTT AGT GAT AAG GAG  186
Allele 2    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --C --- --- --- --->
             -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -

CD1c                                                33
             W   A   R   G   Q   G   S   G   W   L   D   G   L   Q   T   H   G   W   D   S
Allele 1    TGG GCA CGA GGT CAG GGC TCA GGA TGG CTG GAC GAG TTG CAG ACT CAT GGC TGG GAC AGT  126
Allele 2    --- --- --- --- --- --- --- --- --- --- --T --- --- --- --- --- --- --- --- --->
             -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -

CD1d                46
             D   S   D   T   V   R   S   L   K   P   W   S   Q   G   T   F   S   D   Q   Q
Allele 1    GAC TCG GAC ACC GTC CGC TCT CTG AAG CCT TGG TCC CAG GGC ACG TTC AGC GAC CAG CAG  186
Allele 2    --- --- --- T-- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --->
             -   -   -   S   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -

CD1e                                                                        79
             L   K   N   L   Q   S   L   F   Q   L   Y   F   H   S   F   I   Q   I   V   Q
Allele 1    CTG AAA AAC TTA CAG TCA CTG TTC CAG TTA TAC TTC CAT AGT TTT ATC CAG ATA GTG CAA  246
Allele 2    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -G- --- --- --->
             -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   R   -   -   -
```

Fig. 2. **Nucleotide and deduced amino acid substitutions in exon 2 of all five CD1 genes.** Accession numbers for allele 2 of each CD1 subfamily gene are from GenBank: CD1a: AF142665; CD1b: AF142666; CD1c: AF142667; CD1d: AF142668; and CD1e: AF142669.

Allele frequencies for CD1a–CD1e

	Genotype			Allele frequency		% of individuals positive for allele	
n=110	1/1	1/2	2/2	1	2	1	2
CD1a	3	22	85	0.13	0.87	23	98
CD1b	108	2	0	0.99	0.01	100	1
CD1c	98	12	0	0.95	0.04	100	8
CD1d	108	2	0	0.99	0.01	100	2
CD1e	30	41	39	0.49	0.51	69	72

Table 1

formative samples. A summary of the sequence data is shown in Fig. 2 and frequencies for CD1a-CD1e alleles are shown in Table 1. The CD1 sequences present in Genbank were assigned the designation allele 1.

CD1a contained two linked substitutions at amino acid positions 13 and 51. The codon for amino acid 13 has a nucleotide change from T to C resulting in the replacement of an isoleucine with a threonine, I13T. The codon for amino acid 51 has a nucleotide change from G to C resulting in the replacement of a tryptophan with a cysteine, W51C. Allele 1 was present at a frequency of 0.13 and allele 2 at a frequency of 0.87.

Silent substitutions were observed in exon 2 of both CD1b and CD1c (Fig. 1). The newly defined alleles were rare being observed at frequencies of 0.01 and 0.04 respectively (Table 1). Polymorphism in CD1d was also rare. A conservative amino acid substitution at codon 46 was observed in CD1d genes. The nucleotide change of an A to a T would result in an amino acid replacement of Threonine to Serine T46S (Fig. 1). The frequency of allele 2 was 0.01 (Table 1).

CD1e showed two SSCP patterns for exon 2. A nucleotide substitution in codon 79 changed A to G, which would cause an amino acid replacement of glutamine to arginine, Q79R (Fig. 2). The frequency for allele 1 was 0.49 and for allele 2 was 0.51 (Table 1). CD1e SSCP bands segregated as alleles in five families (data not shown).

The crystal structure of murine CD1d was solved recently (5). This structure was used to model the positions of the substitutions in human CD1a and CD1e. The replacement substitutions in exon 2 of CD1a are marked with red spheres in the α1 domain of the CD1 molecule as shown in figure 3A. Both substitutions are in the β pleated sheet portion of the α1 domain; I13T is located above β_2-m and W51C is located in an exposed position on a loop of CD1a on the opposite side of the molecule from the putative CD8 binding site. In contrast, the substitution in CD1e, Q79R, is located in the α-helix region of the α1 domain at the top of the putative ligand-binding groove (Fig. 3B).

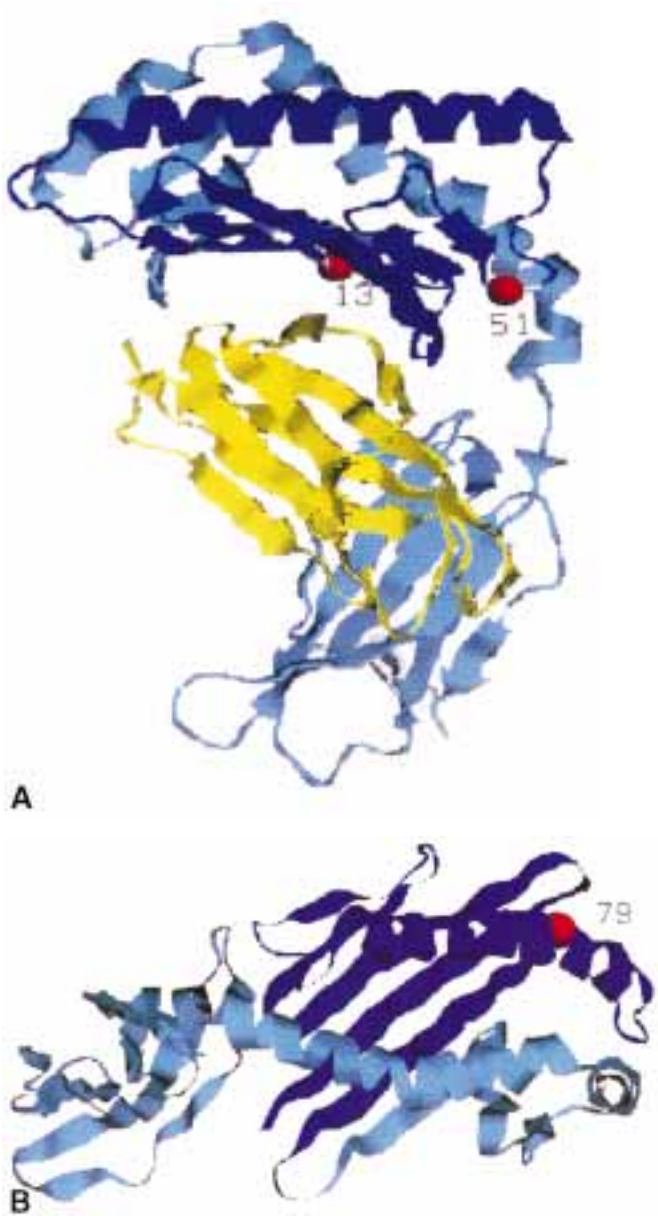

Fig. 3. **Ribbon models of CD1 showing the locations of polymorphic residues in CD1a (A) and CD1e (B).** Polymorphic sites are indicated by red spheres and the amino acid residue number, the α1 domain is colored dark blue, the α2 and α3 domains are colored light blue, and β_2-m is colored yellow. The crystal structure of murine CD1d (5) was used to model structures of human CD1.

Discussion

CD1 genes have been regarded as nonpolymorphic (14–19). By PCR-SSCP analysis, polymorphism in all five human CD1 genes (CD1a, -b, -c, -d, and -e) was detected in samples from a large panel of donors representing diverse ethnic groups. The results showed that

all five CD1 genes are polymorphic. The polymorphisms in CD1a, CD1d, and CD1e sequences resulted in amino acid replacements while polymorphisms in CD1b and CD1c genes were silent. Both alleles of CD1a and CD1e were frequently observed in the population studied whereas the newly identified variants of CD1b, CD1c, and CD1d were infrequent.

The crystal structure of murine CD1d (5) provides a model to locate the positions of substitutions in human CD1a and CD1e. The nature of the replacements and their locations were examined to allow speculation on the functional implications of the substitutions. The two substitutions in CD1a are located in the β pleated sheet region of the α1 domain. Allele 1 contains an isoleucine at position 13 and a tryptophan at position 51 whereas allele 2 has a threonine at 13 and a cysteine at 51.

The replacement of the isoleucine for a threonine at position 13 results in an increase in hydrophobicity and may have an impact upon the interactions with β_2-m. Structural differences have been detected in CD1 heavy chain expressed in the absence of β_2-m. CD1 heavy chains alone are not detected by antibodies but can present antigens to T lymphocytes (20). Both of the residues at position 51, tryptophan or cysteine, are hydrophobic but the cysteine residue has the ability to form disulfide bonds. The potential disulfide linkage could be with other CD1a molecules creating a more stable complex for the association of accessory molecules such as CD8 (21). The region predicted to interact with CD8 is located on the opposite side of the molecule from the cysteine residue (5). Therefore, complexes of CD1a and CD8 could form more stable structures with two molecules of CD1a covalently attached. Alternatively, the cysteine could form a linkage with other cell surface proteins.

In CD1e, the uncharged glutamine at position 79 of allele 1 is changed to a positively charged arginine in allele 2 located in the α helix region flanking the ligand binding groove (Fig. 3B). The change of charge could be significant because the majority of residues that form the ligand binding groove of CD1 molecules are non polar or hydrophobic. This alteration of charge characteristics on the face of the ligand binding groove could impact the specificity of ligands bound and/or their recognition by T lymphocytes. Although CD1e expression has not been demonstrated to date, there are no obvious defects to preclude expression revealed by sequence analyses.

The role of CD1 molecules in antigen presentation has been revealed recently and the fine specificity of T lymphocytes that recognize antigen in the context of CD1 antigens are being elucidated (7–9, 19, 20, 22–29). CD1 molecules have significant similarities to MHC class I molecules in structure and function, yet they also have quite unique characteristics. The tissue distribution of each CD1 antigen is different whereas class I antigens are expressed on all nucleated cells (30). The nature of the ligand binding grooves of class I and CD1 molecules differ and thus peptide and glycolipid or lipid antigens are presented respectively. The extent of genetic polymorphism in class I and CD1 genes is significantly different. Class I genes are highly polymorphic. Although CD1 genes have generally been considered non polymorphic, the present studies have shown that polymorphism is present in all five human CD1 genes. However, only polymorphism in CD1a and CD1e may be functionally significant considering the nature and location of the amino acid substitutions and the frequency of both allelic forms in the human population.

References

1. Calabi F, Millstein C. A novel family of human major histocompatibility complex-related genes not mapping to chromosome 6. *Nature* 1986: **323**: 540–3.
2. Martin LH, Calabi F, Milstein C. Isolation of CD1 genes: A family of major histocompatibility complex-related differentiation antigens. *Proc Natl Acad Sci U S A* 1986: **83**: 9154–8.
3. Yu CY, Mulatein C. A physical map linking the five CD1 human thymocyte differentiation antigen genes. *EMBO* 1989: **8**: 3727–32.
4. Calabi F, Jarvis JM, Martin L, Milstein C. Two classes of CD1 genes. *Eur J Immunol* 1989: **19**: 285–92.
5. Zeng Z, Castano AR, Segelke BW, Stura EA, Peterson PA, Wilson IA. Crystal structure of mouse CD1: An MHC-like fold with a large hydrophobic binding groove [see comments]. *Science* 1997: **277**: 339–45.
6. Rammensee H-G, Falk K, Rotzschke O. Peptides naturally presented by MHC class I molecules. *Annu Rev Immunol* 1993: **11**: 213–44.
7. Brossay L, Chioda M, Burdin N et al. CD1d-mediated recognition of an alpha-galactosylceramide by natural killer T cells is highly conserved through mammalian evolution. *J Exp Med* 1998: **188**: 1521–8.
8. Jackman RM, Moody DB, Porcelli SA. Mechanisms of lipid antigen presentation by CD1. *Crit Rev Immunol* 1999: **19**: 49–63.
9. Spada FM, Koezuka Y, Porcelli SA. CD1d-restricted recognition of synthetic glycolipid antigens by human natural killer T cells. *J Exp Med* 1998: **188**: 1529–34.

10. Rosat JP, Grant EP, Beckman EM et al. CD1-restricted microbial lipid antigen-specific recognition found in the CD8+ alpha beta T cell pool. *J Immunol* 1999: **162**: 366–71.
11. Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc Natl Acad Sci U S A* 1989: **86**: 2766–70.
12. Robinson M, Long E, Johnson A, Hartzman R, Mach B, Kindt T. Recombination within the HLA-D region: correlation of molecular genotyping with functional data. *J Exp Med* 1984: **160**: 222–38.
13. Sayle RA, Milner-White EJ. RASMOL: biomolecular graphics for all. *Trends Biochem Sci* 1995: **20**: 374.
14. van Agthoven A, Terhorst C. Further biochemical characterization of the human thymocyte differentiation antigen T6. *J Immunol* 1982: **128**: 426–32.
15. Amiot M, Dastot H, Fabbi M, Degos L, Bernard A, Boumsell L. Intermolecular complexes between three human CD1 molecules on normal thymus cells. *Immunogenetics* 1988: **27**: 187–95.
16. van de Rijn M, Lerch PG, Knowles RW, Terhorst C. The thmic differentiation markers T6 and M241 are two unusual MHC class I antigens. *J Immunol* 1983: **131**: 851–5.
17. Bendelac A. CD1: Presenting unusual antigens to unusual T lymphocytes. *Science* 1995: **269**: 185–6.
18. Porcelli SA. The CD1 family: a third lineage of antigen-presenting molecules. *Adv Immunol* 1995: **59**: 1–98.
19. Porcelli SA, Segelke BW, Sugita M, Wilson IA, Brenner MB. The CD1 family of lipid antigen-presenting molecules. *Immunol Today* 1998: **19**: 362–8.
20. Brossay L, Burdin N, Tangri S, Kronenberg M. Antigen-presenting function of mouse CD1: one molecule with two different kinds of antigenic ligands. *Immunol Rev* 1998: **163**: 139–50.
21. Blue ML, Levine H, Daley JF, Branton KR, Schlossman SF. Expression of CDI and class I MHC antigens by human thymocytes. *J Immunol* 1989: **142**: 2714–20.
22. Fairhurst RM, Wang CX, Sieling PA, Modlin RL, Braun J. CD1 presents antigens from a gram-negative bacterium, Haemophilus influenzae type B. *Infect Immun* 1998: **66**: 3523–6.
23. Fairhurst RM, Wang CX, Sieling PA, Modlin RL, Braun J. CD1-restricted T cells and resistance to polysaccharide-encapsulated bacteria. *Immunol Today* 1998: **19**: 257–9.
24. Fujieda S, Sieling PA, Modlin RL, Saxon A. CD1-restricted T-cells influence IgG subclass and IgE production. *J Allergy Clin Immunol* 1998: **101**: 545–51.
25. Grant EP, Degano M, Rosat JP et al. Molecular recognition of lipid antigens by T cell receptors. *J Exp Med* 1999: **189**: 195–205.
26. Lewinsohn DM, Alderson MR, Briden AL, Riddell SR, Reed SG, Grabstein KH. Characterization of human CD8+ T cells reactive with Mycobacterium tuberculosis-infected antigen-presenting cells. *J Exp Med* 1998: **187**: 1633–40.
27. Porcelli SA, Brenner MB. Antigen presentation: mixing oil and water. *Curr Biol* 1997: **7**: 508–11.
28. Prigozy TI, Kronenberg M. Presentation of bacterial lipid antigens by CD1 molecules. *Trends Microbiol* 1998: **6**: 454–9.
29. Stenger S, Niazi KR, Modlin RL. Down-regulation of CD1 on antigen-presenting cells by infection with Mycobacterium tuberculosis. *J Immunol* 1998: **161**: 3582–8.
30. Calabi F, Bradbury A. The CD1 system. *Tissue Antigens* 1991: **37**: 1–9.