Functional CD1d and/or NKT cell invariant chain transcript in horse, pig, African elephant and guinea pig, but not in ruminants

Citation for published version:
van Beeck, FAL, Reinink, P, Hermse, R, Zajonc, DM, Laven, MJ, Fun, A, Troskie, M, Schoemaker, NJ, Morar, D, Lenstra, JA, Vervelde, L, Rutten, VPMG, van Eden, W & Van Rhijn, I 2009, 'Functional CD1d and/or NKT cell invariant chain transcript in horse, pig, African elephant and guinea pig, but not in ruminants' Molecular Immunology, vol 46, no. 7, pp. 1424-1431. DOI: 10.1016/j.molimm.2008.12.009

Digital Object Identifier (DOI):
10.1016/j.molimm.2008.12.009

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Molecular Immunology

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Functional CD1d and/or NKT cell invariant chain transcript in horse, pig, African elephant and guinea pig, but not in ruminants

Frank A. Looring van Beecka,1, Peter Reininka,1, Roel Hermsena, Dirk M. Zajoncb, Marielle J. Lavena, Axel Fun, Milana Troskie, Nico J. Schoemakerd, Darshana Morarc, Johannes A. Lenstra, Lonneke Vervelde, Victor P.M.G. Rutten, Willem van Edena, Ildiko Van Rhijna,1,*

a Department of Infectious Diseases and Immunology, Faculty of Veterinary Medicine, University Utrecht, Utrecht, The Netherlands
b Institute of Risk Assessment Science, Utrecht, The Netherlands
c Division of Cell Biology, La Jolla Institute for Allergy and Immunology, La Jolla, CA, USA
d Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria, South Africa
e Department of Clinical Sciences of Companion Animals, University Utrecht, Utrecht, The Netherlands
f Institute of Risk Assessment Science, Utrecht, The Netherlands

1 These authors contributed equally to this manuscript.

ABSTRACT

CD1d-restricted invariant natural killer T cells (NKT cells) have been well characterized in humans and mice, but it is unknown whether they are present in other species. Here we describe the invariant TCR α chain and the full length CD1d transcript of pig and horse. Molecular modeling predicts that porcine (po) invariant TCR α chain/poCD1d/α-GalCer and equine (eq) invariant TCR α chain/eqCD1d/α-GalCer form complexes that are highly homologous to the human complex. Since a prerequisite for the presence of NKT cells is the expression of CD1d protein, we performed searches for CD1D genes and CD1d transcripts in multiple species. Previously, cattle and guinea pig have been suggested to lack CD1D genes. The CD1D genes of European taurine cattle (Bos taurus) are known to be pseudogenes because of disrupting mutations in the start codon and in the donor splice site of the first intron. Here we show that the same mutations are found in six other ruminants: African buffalo, sheep, bushbuck, bongo, N'Dama cattle, and roe deer. In contrast, intact CD1d transcripts were found in guinea pig, African elephant, horse, rabbit, and pig. Despite the discovery of a highly homologous NKT/CD1d system in pig and horse, our data suggest that functional CD1D and CD1d-restricted NKT cells are not universally present in mammals.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

CD1d proteins are expressed on the surface of a variety of antigen presenting cells and non-hematopoietic cells, and present cellular self-lipids and exogenous lipids with an α-anomerically linked sugar to T cells with a highly conserved, invariant TCR, NKT cells. CD1d−/− mice have no detectable mature NKT cells (Chen et al., 1997; Gapin et al., 2001), showing that functional CD1D genes are a prerequisite for their development. NKT cells have been implicated in oral tolerance, autoimmunity, dendritic cell maturation, tumor surveillance, and anti-microbial immunity. Natural exogenous ligands for NKT cells presented by CD1d include highly pathogenic bacteria, peptidoglycan and LPS-negative bacteria, including pathogenic tick-borne genera: Rickettsia, Anaplasma, and Ehrlichia, all causing morbidity and mortality in livestock. Unfortunately, none of these bacteria has been studied closely enough to determine whether they contain antigens for NKT cells. Ehrlichia ruminantium (formerly named Cowdria ruminantium (Dumler et al., 2001)), causes heartwater (cowdriosis), Anaplasma bovis (formerly named Ehrlichia bovis) causes bovine ehrlichiosis, and A. marginale and A. centrale cause bovine anaplasmosis, and these diseases are major problems in the livestock industry in sub-Saharan Africa. Some indigenous African breeds of cattle are more resistant to heartwater and anaplasmosis than other...
breeds, but this can be explained by higher resistance to the vector (ticks of the genus *Amblyomma*). All breeds of cattle will develop clinical disease once they get infected. Since the aforementioned bacterial pathogens do not carry the signature danger molecules LPS and peptidoglycan, recognition by the innate immune system other than TLRs, like the CD1d/NKT system, may be of crucial importance in the early defense against these pathogens.

It has been suggested that the group 1 CD1 proteins (CD1a, CD1b, CD1c) are not universally present in all species, whereas group 2 CD1 proteins (CD1d) are. CD1D genes have indeed been found in many mammalian species studied, including primates like humans and chimpanzees (*Pan troglodytes*), African green monkeys (*Chlorocebus sabaeus*) and rhesus macaques (*Macaca mulatta*) (Saito et al., 2005), mice (*Mus musculus*) (Bradbury et al., 1988), rats (*Rattus norvegicus*) (Ichimiya et al., 1994), cottontail rabbits (*Sylvilagus floridanus*) (Calabi et al., 1989), sheep (*Ovis aries*) (Rhind et al., 1999), and pigs (*Sus scrofa*) (Eguchi-Ogawa et al., 2007). However, not all of these genes have been shown to lead to functional transcripts or proteins yet. CD1 genes have also been discovered in chickens (*Gallus gallus*) (Maruoka et al., 2005; Miller et al., 2005; Salomonsen et al., 2005), but chicken CD1 genes could not be classified according to the existing isoforms, and are therefore named CD1.1 and CD1.2. There are two species that have till now been suggested to have no functional CD1D genes. Before the availability of its genome, the guinea pig (*Cavia porcellus*) family of CD1 genes had been well characterized, but a CD1D gene was not identified (Dascher et al., 1999). In cattle, two CD1D genes have been identified, named CD1D1 and CD1D2, but these are in fact pseudogenes (Van Rhijn et al., 2006). The two CD1D pseudogenes that were identified both contain a mutated start codon and an unspliecex intron. In this paper we describe CD1D pseudogenes in N'Dama cattle, and five other ruminants, including sheep, which had previously been assumed to have functional CD1D genes. Functional CD1D transcripts were identified in guinea pig, African elephant, horse, rabbit, and pig.

NKT cells can be distinguished by their highly conserved invariant TCR. The NKT cell population can be visualized by flow cytometric analysis using fluorescently labeled CD1d tetramers and/or NKT cells are present in several mammalian species, like dog and horse as described in this paper, are likely to be true functional homologs. Like dog and horse, NKT cells can be distinguished by their highly conserved invariant TCR. The NKT cell population can be visualized by flow cytometric analysis using fluorescently labeled CD1d tetramers and/or NKT cells are present in several mammalian species, like dog and horse as described in this paper, are likely to be true functional homologs.

Our data provide supportive evidence that functional CD1d transcripts and/or NKT cells are present in several mammalian species, but not in ruminants. This shows that the CD1D/NKT system is not universally present as previously thought. The CD1D/NKT system may be lacking in ruminants altogether, providing a possible explanation for their high sensitivity to *Rickettsia*, *Anaplasma*, and *Ehrlichia*.

### 2. Materials and methods

#### 2.1. Animals and nucleic acid samples

Genomic DNA from sheep (*Ovis aries*), African buffalo (*Syncerus caffer*), bushbuck (*Tragelaphus scriptus*), and bongo (*Tragelaphus eurycerus*), was freshly isolated from, heparrinized blood, using the Wizard Genomic DNA Purification Kit (Promega). Genomic DNA from N'Dama cattle (*Bos taurus*) and roe deer (*Capreolus capreolus*) were isolated at least 10 years ago. DNA from N'Dama cattle was a gift of Dr. Bradley (Trinity College, Dublin, Ireland). First strand cDNA was synthesized with Multiscribe reverse transcriptase (Applied Biosystems) from RNA isolated with the Qiagen RNAEasy kit from freshly isolated PBMC from cattle, sheep, pig (*Sus scrofa*), horse (*Equus caballus*), guinea pig (*Cavia porcellus*), cat (*Felis catus*), rabbit (*Oryctolagus cuniculus*), and African elephant (*Loxodonta africana*). Animal-derived materials were obtained according to the regulations of the Animal Ethical Committee of the University Utrecht, The Netherlands (protocol number 2007.11.06.152/Vervolg1).

#### 2.2. Searches in genomes and EST databases

BLAST searches were performed in selected genomes (www.ensembl.org) and EST databases (www.ncbi.nlm.nih.gov/BLAST) with the nucleotide sequence of the α1 and the α2 domains of human CD1D (*NM_001766*) and with the human TRAV10 segment (also called Vε24) used by NKT cells (*AE000659*). The results of the CD1D searches were included in a phylogenetic tree together with the α1 and the α2 domains of the known CD1 isoforms to assess with which CD1 isoforms they group. Also, the obtained potential CD1D genes were checked for the presence of a leader peptide, α1, α2, and α3 domain, and a transmembrane region. The nucleotide sequence of the α1 domain of the published sheep CD1D cDNA (*AJ006722*) was used to perform a BLAST search in the NCBI sheep EST database (www.ncbi.nlm.nih.gov/BLAST). The predicted amino acid sequences of the hits obtained from BLAST searches with the human TRAV10 segment in the genomes of selected species were all aligned and evaluated as described in Section 3.

#### 2.3. Primer design, PCR and sequencing

Universal primers CD1D1 primers (Invitrogen) were designed based on the sequences of bovine pseudogene CD1D1 (accession number *DQ192544*). Forward primers “K” (5′-GGCCGAGATTCGAGTCGCC-3′) and “S” (5′-CGTGTTGGGGAGGACGGAGG-3′) ended approximately 20 and 50 base pairs, respectively upstream of the start codons. The reverse primers “N” (5′-CTTCTCTCAGCTCMGACTTCC-3′) and “M” (5′-CAGCCGCACTGGGCCCACCTG-3′) were designed to anneal to the end of the α2 domain and the beginning of the α3 domain, respectively. Primers for the full length CD1D transcripts of guinea pig, African elephant, rabbit, and horse were designed based on the predicted sequences in their genomes: CporCD1DFor#1 (5′-TGTCAGAAGTCCGGGTCC-3′), CporCD1D-Rev#2 (5′-AACTTTGAGTACCAAGATCTTTGAG-3′), LafriCD1DFor#1 (5′-GAGCCGAGAGTTGGAGGAGG-3′), and OcunCD1DFor#2 (5′-CTGGCAGCTGCCTGTGAG-3′), OcunCD1DRev#1 (5′-CTACAAGATGGCTGTGAG-3′), EbcdCD1DFor#1 (5′-CTCAACCCGACGGAGAGAG-3′), EbcdCD1DRev#1 (5′-CTGACAGATTGGAGGAGGAGG-3′), FcatCD1DRev#1 (5′-CTTCTCTGAGTACCAAGAAC-3′), CporTCRαFor#1 (5′-CTTCTCTGAGTACCAAGAAC-3′), CporTCRαRev#1 (5′-CTTCTCTGAGTACCAAGAAC-3′), and CporTCRαRev#1 (5′-CTTCTCTGAGTACCAAGAAC-3′). The sequences were amplified using a V segment-specific forward primer and a reverse primer in the constant segment: CporTCRαFor#1 (5′-CTTCTCTGAGTACCAAGAAC-3′), CporTCRαRev#1 (5′-CTTCTCTGAGTACCAAGAAC-3′), CporTCRαRev#1 (5′-GAGCCGAGAGTTGGAGGAGG-3′), CporTCRαRev#1 (5′-GAGCCGAGAGTTGGAGGAGG-3′), CporTCRαRev#1 (5′-CTCAACCCGACGGAGAGAG-3′), CporTCRαRev#1 (5′-CTCAACCCGACGGAGAGAG-3′), CporTCRαRev#1 (5′-CTCAACCCGACGGAGAGAG-3′), and CporTCRαRev#1 (5′-GAGCCGAGAGTTGGAGGAGG-3′).
AGG TTT AGG TTG TTC-3′), OariTCrRevFor#1 (5′-CAA-
TTTACAGTGACCCCCCTAAACA-3′), OariTCrRev#1 (5′-GAAC-
ACGGTGACTCAGGTTTTG-3′), BtauTCrFor (5′-CAATGGACGT-
ACACAGCCAG-3′), BtauTCrRev “J” (5′-TTTCAAGGTTTTC-
TCTACAGGTTTG-3′). PCR was performed with PFU Turbo poly-
merase (Stratagene) according to the protocol of the manufacturer
using 20 ng genomic DNA or 10 ng cDNA in a reaction volume of
20 μl under the following cycling conditions: an initial denatura-
tion of 7 min at 95 °C, followed by 40 cycles of 30 s at 95 °C, 45 s
at different annealing temperatures (60 °C, 57 °C, 54.5 °C), 1 min
at 72 °C, followed by a final elongation step of 7 min at 72 °C. PCR
products were cut from a 1.5% agarose gel, purified with Gel DNA
recovery kit (2ymo), and ligated in a Topo4Blunt vector which
was used to transform one shot Top10 cells (Invitrogen). DNA
inserts of single colonies was sequenced by Baseclear (Leiden, The
Netherlands).

2.4. Sequence analysis and homology modeling

Homology models of pig (Sus scrofa) and horse (Equus caballus)
CD1d, as well their α chain of the invariant NKT cell TCR were mod-
eled using the Swiss Model Server (Schwede et al., 2003), using both
human CD1d and Vα24 TCR crystal structures as templates. The obtained
CD1d and TCR models were superimposed onto their corre-
csponding human counterparts in the CD1d/α-GalCer/Vα24 TCR
crystal structure (PDB code 2PO6) (Borg et al., 2007). No reorienta-
tion of the TCR was necessary to accommodate the TCR CDR loops,
due to their similar orientation in both models. The CD1d surface
residues in all three CD1d orthologs are mostly conserved, except
for a glycine residue instead of the human tryptophan (W153),
which is responsible for tilting the galactose of α-GalCer when
bound to human CD1d (Koch et al., 2005) in comparison to mouse
CD1d. Therefore, we manually modeled this galactose in the or-
testation that it adopts when bound to mouse CD1d, as mCD1d also
has this conserved glycine residue (Zajonc et al., 2005). The models
were visualized using PyMol (pymol.sourceforge.net).

The Translate Nucleic Acid Sequence Tool was used (http://
bioinformatics.umd.edu) for translation into amino acids. Align-
ments were performed and trees generated with ClustalW
biotools.umassmed.edu) for translation into amino acids. Align-
ments were performed and trees generated with ClustalW

3. Results

3.1. Invariant α chain analysis

The human TRAV10V segment (also called Vα24) that is used
by the human NKT invariant TCR was used to identify TCR α chain
V segments in the genomes of cat, dog, horse, pig, cat, guinea
pig, African elephant, rabbit, and sheep. All resulting V α segment
sequences were translated and aligned with the human TRAV10
segment. We considered all V segments with higher sequence
homology to TRAV10 than to any other human V segment as can-
didate V segments for the NKT invariant chain in other species.
Because the CDR1 region is encoded by the V segment and known to
interact with α-GalCer (Borg et al., 2007; Kjer-Nielsen et al., 2006;
Scott-Browne et al., 2007), we only included V segments in which
at least two residues, including the P that was indicated as crucial
in all studies, were identical to the human TRAV10 CDR1 region
(VSFP5N). According to these criteria we identified one candidate V
segment in cat, dog, horse, pig, guinea pig, African elephant, rabbit,
and sheep, and three in cattle (Table 1a).

Using a forward primer before or at the CDR1 region of the can-
didate V segments and a reverse primer in the constant segment,
we amplified partial TCR α chains covering the CDR1, CDR2, CDR3,
and part of the constant domain. For this purpose, PBMC-derived
cDNA was available from cat, dog, horse, pig, guinea pig, African ele-
phant, rabbit, sheep and cattle. CDR3 sequences that were highly
homologous to the human and murine NKT CDR3α were obtained
from horse (two out of four sequences) and pig (one out of 11
sequences). Six out of eight sequences obtained from cat had a
two amino acid deletion in the CDR3 compared to the human and
murine sequences. From cattle, one out of 15 sequences showed
high homology to the human CDR3, but it had one extra amino
acid. None of 15 sheep sequences, eight guinea pig sequences,
eight rabbit sequences, and one African elephant sequence showed
homology to the human invariant CDR3α (Table 1b). We were
not able to derive TCR α chain sequences from dog. To predict
whether the obtained CDR1α and CDR3α loops would be able to
interact with a CD1d/α-GalCer complex, we generated models
using the Swiss Model server (Schwede et al., 2003), and com-
pared these to the available human data (Borg et al., 2007). The horse
and pig invariant TCR α chain/CD1d/α-GalCer models suggest that
these α chain sequences are fully functional invariant NKT cell
TCR sequences, capable of binding α-GalCer, when presented by its
species-matched CD1d molecule (Fig. 1). Even though otherwise
highly conserved, the differences in CDR3α length of the obtained
bovine and feline sequences make it difficult to predict whether the
residues that normally interact with α-GalCer do so in these species,
and therefore we cannot conclude that these sequences represent
the bovine or feline NKT invariant chain.

3.2. CD1D pseudogenes in ruminants

PCR products were generated using genomic DNA from N'Dama
cattle (Bos taurus), African buffalo (Syncerus caffer), sheep (Ovis
aries), roe deer (Capreolus capreolus), bushbuck (Tragelaphus script-
us), and bongo (Tragelaphus eurycerus), using heterologous CD1
primers. Subsequent cloning of PCR products and sequencing of
at least four independent bacterial colonies of each species
resulted in CD1D sequences available at Genbank with accession
numbers EU247610–EU247617 and FJ028651–FJ028652. In case of
small differences between sequences derived from one species, the
sequence that was closest to the consensus sequence was submitted
to Genbank. Alignment of the newly derived ruminant sequences
with previously published CD1D sequences of humans and cattle
resulted in CD1D1 pseudogenes for the ruminants (Fig. 2a)
revealed that all newly derived ruminant CD1D1 sequences
have the same disrupting mutations as the bovine CD1D1 genes.
The start codon is mutated and the donor splice site of the first intron
(the intron after the leading fragment) is mutated, rendering it an
unspliceable intron. Interestingly, the mutated donor splice site of
that intron forms ATG in all ruminant CD1D1 genes, and might func-
tion as an alternative start codon. This ATG is in the right reading
frame and does in most cases not lead to any premature stop codons.
However, the protein that would be synthesized is not predicted to
contain a leading fragment by the SignalP program and can thus not
be expressed at the cell surface (Signal peptide probability: 0.001,
Signal anchor probability: 0.000). In N’Dama cattle and bongo we
obtained CD1D1 and CD1D2 pseudogenes. Comparison of exons 1–3 of these two sequences, revealed
that they were >98% identical at nucleotide level, suggesting that
FJ028652 may be a transcript of the gene we report here. To obtain
additional data on the status of the sheep CD1D gene, we investi-
gated CD1D transcripts in the sheep EST database. A BLAST search

The published sheep CD1d mRNA sequence with accession
number AJ006722 (Rhind et al., 1999) does not show disruptive
mutations, while the sheep CD1D pseudogene we describe here
does. Comparison of exons 1–3 of these two sequences, revealed
that they were >98% identical at nucleotide level, suggesting that
AJ006722 may be a transcript of the gene we report here. To obtain
additional data on the status of the sheep CD1D gene, we investi-
gated CD1D transcripts in the sheep EST database. A BLAST search

The translated sequences were obtained using the Swiss Model
server (http://swissmodel.expasy.org) for model generation.

The published sheep CD1d mRNA sequence with accession
number AJ006722 (Rhind et al., 1999) does not show disruptive
mutations, while the sheep CD1D pseudogene we describe here
does. Comparison of exons 1–3 of these two sequences, revealed
that they were >98% identical at nucleotide level, suggesting that
AJ006722 may be a transcript of the gene we report here. To obtain
additional data on the status of the sheep CD1D gene, we investi-
gated CD1D transcripts in the sheep EST database. A BLAST search

The published sheep CD1d mRNA sequence with accession
number AJ006722 (Rhind et al., 1999) does not show disruptive
mutations, while the sheep CD1D pseudogene we describe here
does. Comparison of exons 1–3 of these two sequences, revealed
that they were >98% identical at nucleotide level, suggesting that
AJ006722 may be a transcript of the gene we report here. To obtain
additional data on the status of the sheep CD1D gene, we investi-
gated CD1D transcripts in the sheep EST database. A BLAST search

The published sheep CD1d mRNA sequence with accession
number AJ006722 (Rhind et al., 1999) does not show disruptive
mutations, while the sheep CD1D pseudogene we describe here
does. Comparison of exons 1–3 of these two sequences, revealed
that they were >98% identical at nucleotide level, suggesting that
AJ006722 may be a transcript of the gene we report here. To obtain
additional data on the status of the sheep CD1D gene, we investi-
gated CD1D transcripts in the sheep EST database. A BLAST search

The published sheep CD1d mRNA sequence with accession
number AJ006722 (Rhind et al., 1999) does not show disruptive
mutations, while the sheep CD1D pseudogene we describe here
does. Comparison of exons 1–3 of these two sequences, revealed
that they were >98% identical at nucleotide level, suggesting that
AJ006722 may be a transcript of the gene we report here. To obtain
additional data on the status of the sheep CD1D gene, we investi-
gated CD1D transcripts in the sheep EST database. A BLAST search
Table 1
Sequences of V segments homologous to TRAV10 and CDR3 of T cells using these V segments.

### a) V segments homologous to hu Vα24

| Species | V segment sequence |
|---------|--------------------|
| Human   | GKNTLQCNYTETGLRWFYQDTOGPGPVSLITMTFSEKTRSGRYTAGALTDA7KS8LH1TAQSLDSASAYICV |
| Mouse   | GENVTCQNYSLRWFYQDTOGPGPVSLITMTFSEKTRSGRYTAGALTDA7KS8LH1TAQSLDSASAYICV |
| Afr. elephant | GENVTCQNYSLRWFYQDTOGPGPVSLITMTFSEKTRSGRYTAGALTDA7KS8LH1TAQSLDSASAYICV |
| Guinea pig | GSREGTdCNYTETGLRWFYQGPGPGLYTVLASSK5K7SGRTGYTVLVDVAHKSSEL1TAQS5Q5DSASAYICV |
| Pig     | GENTIHCNTYELRWFYQDTOGPGPVSLITMTFSEKTRSGRYTAGALTDA7KS8LH1TAQSLDSASAYICV |
| Sheep   | GENTIHCNTYELRWFYQDTOGPGPVSLITMTFSEKTRSGRYTAGALTDA7KS8LH1TAQSLDSASAYICV |
| Dog     | GENTIHCNTYELRWFYQDTOGPGPVSLITMTFSEKTRSGRYTAGALTDA7KS8LH1TAQSLDSASAYICV |
| Cat     | GENTIHCNTYELRWFYQDTOGPGPVSLITMTFSEKTRSGRYTAGALTDA7KS8LH1TAQSLDSASAYICV |
| Horse   | GENTIHCNTYELRWFYQDTOGPGPVSLITMTFSEKTRSGRYTAGALTDA7KS8LH1TAQSLDSASAYICV |
| Rabbit  | GENTIHCNTYELRWFYQDTOGPGPVSLITMTFSEKTRSGRYTAGALTDA7KS8LH1TAQSLDSASAYICV |
| Cattle  | GENTIHCNTYELRWFYQDTOGPGPVSLITMTFSEKTRSGRYTAGALTDA7KS8LH1TAQSLDSASAYICV |

### b) CDR3α with high homology to human and mouse invariant CDR3α

| Species | Clone name | CDR3 sequence | Frequency |
|---------|------------|---------------|-----------|
| Human   | CVSESRGLGRLGY |            |           |
| Mouse   | CVGDSGSLGRLGY |            |           |
| Pig     | susc6       | CVGDSGSLGRLGY | 1x         |
| Horse   | clone1      | CVGDSGSLGRLGY | 2x         |
| Cat     | feliscat2   | CVV--RSRGLLGY | 6x         |
| Cattle  | PRK11       | CVGDSGSLGRLGY | 1x         |

### CDR3α with low homology to human and mouse invariant CDR3α

| Species | Clone name | CDR3 sequence | Frequency |
|---------|------------|---------------|-----------|
| Pig     | susc1      | CVGVRGSGSGSLTGL | 1x         |
|         | susc3      | CVGVGRGSGSLTGL | 1x         |
|         | susc4      | CVGVGRGSGSLTGL | 1x         |
|         | susc7      | CVGVGRGSGSLTGL | 1x         |
|         | susc10     | CVGVGRGSGSLTGL | 1x         |
|         | susc16     | CVGVGRGSGSLTGL | 2x         |
|         | susc10     | CVGVGRGSGSLTGL | 1x         |
| Guinea pig | c1       | CVGVGRGSGSLTGL | 1x         |
|         | c8         | CVGVGRGSGSLTGL | 1x         |
|         | c9         | CVGVGRGSGSLTGL | 1x         |
|         | c10        | CVGVGRGSGSLTGL | 1x         |
|         | C9         | CVGVGRGSGSLTGL | 1x         |
| Cat     | feliscat6  | CVGVGRGSGSLTGL | 1x         |
| Rabbit  | PRE11      | CVGVGRGSGSLTGL | 1x         |
|         | rab7       | CVGVGRGSGSLTGL | 1x         |
|         | rab9       | CVGVGRGSGSLTGL | 1x         |
|         | rab11      | CVGVGRGSGSLTGL | 1x         |
|         | PRU1       | CVGVGRGSGSLTGL | 1x         |
|         | PRU2       | CVGVGRGSGSLTGL | 1x         |
|         | PRU3       | CVGVGRGSGSLTGL | 1x         |
|         | PRU7       | CVGVGRGSGSLTGL | 1x         |
|         | c18        | CVGVGRGSGSLTGL | 2x         |

a Amino acid sequences of TRAV10-homologous V segments in several species, identified by searching the available genomic data. In green: CDR1. b TCR α chain sequences using the TRAV10 homologs were derived from PBMC from multiple species. The CDR3 of these TCR α chains that are highly homologous to the CDR3 of the human and mouse NKT TCR α chain are aligned (top panel). The human and mouse sequences that are included were derived from literature. CDR3 that were not homologous to the CDR3 of the human and mouse NKT TCR α chain, but were used by TRA10 homologous V segments are shown for comparison (lower panel). Green: CDR1; Grey: the first two amino acids of the FCGXG motif, forming the end of the CDR3.
with the nucleotide sequence of exon 2, encoding the α1 domain of the AJ006722 sequence resulted in five hits that were >98% identical at nucleotide level, suggesting that they were transcripts of the same gene. Three of these hits (EE803429, DY491833, and DY491595) contained a mutated start codon and an unspliceable intron between the leading fragment and the α1 domain. The other two hits did not contain any sequence upstream of the α1 domain. From this we conclude that in the EST database there are no functional CD1D transcripts corresponding to the AJ006722 sequence, but there are transcripts of the pseudogene we describe in this paper. The only sheep CD1 proteins that have been demonstrated at protein level were CD1b and CD1e, isolated by immunoprecipitation with an antibody that recognizes multiple ruminant CD1 molecules (Rhind et al., 1999).

3.3. CD1D genes and CD1d transcripts in non-ruminant species

CD1D sequences were identified in the genomes of dog, cat, pig, guinea pig, horse, African elephant, rabbit, nine-banded armadillo, small Madagascar hedgehog, European shrew, and northern tree shrew (Table 2). A full length CD1D sequence without any of the characteristics of pseudogenes could be found in pig, horse, and nine-banded armadillo. The CD1D sequences of the other mammals were incomplete because of gaps in the genomic sequences. However, the available parts of the sequences did not show any of the characteristics of pseudogenes. In order to obtain the full length coding sequence of the incomplete genes, and proof that the CD1D genes are transcribed and properly spliced in vivo, we successfully cloned full-length CD1d transcripts from guinea pig, rabbit, horse, and African elephant PBMC (accession numbers FJ028653–FJ028656). Alignment of these sequences with the human and murine CD1d sequences (Fig. 3) shows that the residues on the surface of CD1d that interact with the NKT TCR are highly conserved. Contrary to all other CD1D sequences the African elephant CD1d sequence has a truncated cytoplasmic tail and lacks a YXXZ motif. The YXXZ motif in the tail sequence of murine and human CD1d is needed for interaction with AP-2 and thus trafficking to the late endosome (Chiu et al., 1999; Rodionov et al., 1999).

4. Discussion

In this paper we show that the NKT/CD1d system is present in horse and pig. Equine and porcine NKT invariant α chains and CD1d transcripts are sequenced and their models suggest that they are likely to function as their human and murine counterparts. In addition, we sequenced full length CD1d cDNA of African elephant, guinea pig, and rabbit, and we show that in the genomes of dog, cat, African elephant, nine-banded armadillo, Madagascar hedgehog, European shrew, and Northern tree shrew (partial) functional CD1D genes are present, suggesting that these species may also have a functional NKT/CD1d system. However, in the six ruminant species we studied here, all CD1D genes we identified were non-functional, which strongly suggests that ruminants may not have NKT cells.

Table 2

| Species                              | NCBI accession or ensemble gene ID | Link                                      | Functionality |
|--------------------------------------|-----------------------------------|-------------------------------------------|---------------|
| Dog (Canis familiaris)               | AC183576                          | http://www.ncbi.nlm.nih.gov/sites/Nucleotide | Incompletea   |
| Cat (Felis catus)                    | ENSFCAG00000012532                | http://www.ensembl.org/Felis_catus        | Incompletea   |
| Pig (Sus scrofa)                     | NM_001028680.1                    | http://www.ncbi.nlm.nih.gov/sites/Nucleotide | Functionalb   |
| Guinea pig (Cavia porcellus)         | ENSCPOG00000013837                | http://www.ensembl.org/Cavia_porcellus   | Functionalb   |
| Horse (Equus caballus)               | NW_001080880.1                    | http://pre.ensembl.org/Equus_caballus    | Functionalb   |
| African elephant (Loxodonta africana) | ENSLAFG00000013716                | http://www.ensembl.org/Loxodonta_africana | Functionalb   |
| Nine-banded armadillo (Dasypus novemcinctus) | ENSDNOG00000018694        | http://www.ensembl.org/Dasypus_novemcinctus | Functionalb   |
| Madagascar hedgehog (Echinops telfairi) | ENSETEG00000015412               | http://www.ensembl.org/Echinops_telfairi | Incompletea   |
| European shrew (Sorex araneus)       | ENSSARG00000006647                | http://www.ensembl.org/Sorex_araneus     | Incompletea   |
| Northern tree shrew (Tupaia belangeri)| ENSTBEG00000001753                | http://www.ensembl.org/Tupaia_belangeri  | Incompletea   |

a The genomic sequence contained gaps. The available part does not contain any of the characteristics of a pseudogene.

b Full-length transcripts of this gene that are predicted to translate into a functional protein have been described in this paper.

c The gene is complete and did not contain any of the characteristics of a pseudogene, but it is unknown whether the gene is transcribed and translated in vivo.
Human and murine CD1d-restricted NKT cells can be detected using α-GaICer-loaded CD1d tetramers. Even though human and murine CD1d tetramers cross-react between these two species, lack of detection of NKT cells in ruminants using murine or human CD1d tetramers is not conclusive because these reagents may not cross-react with ruminants. Using the same molecular approach that led to the identification of the invariant NKT α chain in pig and horse we were unable to identify invariant CD1d chain homologs in guinea pig, cat, rabbit, African elephant, cattle and sheep. So, even though these species do express V segments homologous to TRAV10 (Vα24), we have not found these V segments in combination with the canonical NKT CDR3. It is possible that we did not obtain invariant NKT α chain sequences because the TRAV10-homologous V segment is used often by other, non-NKT cells in these species, our sample size is not big enough, and/or the NKT cells are underrepresented in PBMC. Therefore, based on TCR α chain sequences only, we cannot conclude that NKT cells are absent in these species. However, the combination of the fact that ruminants lack functional CD1d genes, and the observed absence of an invariant NKT α chain sequence among 26 different ruminant (bovine and ovine) TRAV10-homolog-containing TCR α chain sequences points strongly to absence of NKT cells in ruminants.

CD1d presents lipids with an α-GaICer glycosidic linkage to NKT cells and may therefore be an important molecule to stimulate the
immune system in response to α-proteobacteria that contain these compounds. Ruminants are very sensitive to infection with these pathogens. Previously we have shown that European cattle lack functional genes for CD1D. Because we found CD1D pseudogenes and no functional CD1D genes in African N'Dama cattle (Bos taurus), two other species of the family Bovinae (bongo, bushbuck, and African buffalo), a member of the superfamily Bovidae that does not belong to the family of Bovinae (sheep), and a ruminant that is a member of the superfamily Cervidae (roe deer), we conclude that CD1d proteins are probably absent in all ruminants, though we are aware that we have not formally proven this. In the absence of a fully finished and assembled genome, it is difficult to prove that a functional CD1D gene is absent in a certain species. Southern blotting detects hybridizing sequences, but does not discriminate between functional genes and pseudogenes and between comigrating restriction fragments. Especially the latter poses problems because homology of CD1 genes can be exceptionally high, and these genes will be cut in an identical way by the restriction enzymes used, leading to an underestimation of the real number of CD1 genes. This probably explains that the guinea pig was previously suspected of having no CD1D gene (Dascher et al., 1999) based on Southern blot data, while we report a guinea pig CD1D gene and transcript here.

Our data on the presence of a CD1D pseudogene in sheep, carrying a mutated start codon and an unspliceable intron, seem to be in contradiction with published data on sheep CD1D. A full-length cDNA sequence of sheep CD1D has been published and is predicted to translate into a normal CD1D protein (Rhind et al., 1999). However, this cDNA sequence has been assembled in silico from partial PCR products. The full-length cDNA sequence in which the first intron was properly spliced out has never been obtained (Rhind, personal communication). In the sheep EST database we could only find CD1D pseudogene transcripts, and no functional transcript analogous to the published one. Together, this suggests that the published sheep CD1d cDNA may derive from transcripts of a sheep CD1D pseudogene, and are consistent with the possibility that sheep do not have functional CD1D.

The artiodactyl pig (S. scrofa) is the closest relative of the ruminants that we studied, and it has a functional gene and no CD1D pseudogene. This dates the loss of a functional CD1D gene by point mutations, and thus the emergence of a CD1D pseudogene approximately 65 million years ago, when the ancestors of Suidae and Ruminantia diverged (Kumar and Hedges, 1998). This is consistent with the fact that we have only found CD1D pseudogenes in all ruminants studied and argues against the presence of a functional sheep CD1D gene.

To emphasize the special status of group 2 CD1 molecules as compared to group 1 CD1 molecules, it has often been stated that CD1d molecules and NKT cells are universally present in mammals, while this is not the case for group 1 CD1 molecules. Lack of functional CD1D genes in a considerable group of animals as we show here would suggest that there is no reason for a special status for CD1D proteins based on universal distribution among mammals. In addition to different expression patterns and being slightly separated based on sequence homology, group 2 CD1 molecules (CD1d) are thought to differ fundamentally from group 1 CD1 molecules.
(CD1a, CD1b, and CD1c) in that they stimulate an invariant T cell population. However, in addition to being able to activate NKT cells with an invariant TCR, it has been shown that CD1d can also stimulate other, non-invariant T cells (Baron et al., 2002; Huber et al., 2003; Jahng et al., 2004; Van Rijn et al., 2004). Whether CD1d is the only member of the CD1 family of proteins that can stimulate an invariant T cell population, remains open: it is possible that in the future invariant group 1 CD1-restricted T cell populations will be discovered, and if so, this would question whether group 1 and group 2 CD1 proteins really perform fundamentally different functions.

Acknowledgements

Ildiko Van Rijn is supported by the Wellcome trust [GR078283]. Dirk M. Zajonc is a recipient of a CRI investigator award. We would like to thank Jackie Picard and Basil Allsopp (University of Pretoria, Onderstepoort, South Africa) for expert advice, Dan Bradley (Trinity College, Dublin, Ireland), Astrid Rijkenhuizen, and Ton Willemse (University Utrecht, Utrecht, The Netherlands) for kindly sharing DNA and/or tissue samples.

References

Baron, J.L., Gardiner, L., Nishimura, S., Shinaki, K., Locksley, R., Ganem, D., 2002. Acti-

Gapin, L., Matsuda, J., Surh, C.D., Kronenberg, M., 2001. NKT cells derive from double-

Dascher, C.C., Hiromatsu, K., Naylor, J.W., Brauer, P.P., Brown, K.A., Storey, J.R.,

Chen, Y.H., Chiu, N.M., Mandal, M., Wang, N., Wang, C.R., 1997. Impaired NK1+ T

Calabi, F., Belt, K.T., Yu, C.Y., Bradbury, A., Mandy, W.J., Milstein, C., 1989. The rabbit

Bradbury, A., Belt, K.T., Neri, T.M., Milstein, C., Calabi, F., 1988. Mouse CD1 is distinct

References

College, Dublin, Ireland), Astrid Rijkenhuizen, and Ton Willemse (University Onder-

Baron, J.L., Gardiner, L., Nishimura, S., Shinaki, K., Locksley, R., Ganem, D., 2002. Acti-

Gapin, L., Matsuda, J., Surh, C.D., Kronenberg, M., 2001. NKT cells derive from double-

References