Molecular Characterisation Of Ovine CD1.

Elaine D. Ferguson.

This thesis is submitted as part of the course requirements for the degree of Doctor of Philosophy at the University of Edinburgh.
November, 1994.

Dedication

In memory of my Dad, James Ferguson (1930-1983).

UNIVERSITY OF EDINBURGH

ABSTRACT OF THESIS (Regulation 3.5.10)

Name of Candidate ELAINE DOROTHEA FERGUSON

Address ....

Degree Ph.D. *Date* 3 November 1994

Title of Thesis Molecular Characterisation of Ovine CD1

No. of words in the main text of Thesis 45,085

The CD1 molecules are a family of $\beta_2$microglobulin-associated glycoproteins with strong structural homology, but weaker sequence homology, to the MHC class I antigens. In contrast to the classical class I antigens, CD1 molecules exhibit restricted tissue expression (cortical thymocytes, dendritic cells, a subset of B cells and some intestinal epithelial cells), and are nonpolymorphic. Five CD1 genes have been identified in humans, two in the mouse and several in other mammalian species (Calabi et al, 1991). CD1 expression has also been detected by immunohistological techniques in the cow, sheep and pig.

The MHC class I-like structure of CD1 and the expression on classical antigen presenting cells of the immune system has pointed to a role for CD1 in antigen presentation. Indeed, evidence has been accumulating over the past few years to support this view, with several reports suggesting that $CD4^-8^-$ T cells in particular may be able to recognise nonclassical presentational elements including MHC class Ib molecules such as TLa and Qa, as well as CD1. Most recently, CD1b molecules on human monocytes have been demonstrated to restrict the response of $CD4^-8^-$ T cells to antigens derived from *M. tuberculosis* (Porcelli et al, 1992).

Previous studies on the ovine CD1 family have involved the use of monoclonal antibodies to assess tissue expression and distribution, and biochemical analyses of the ovine CD1 antigens. However, no studies have been carried out to investigate ovine CD1 at the molecular level. Therefore, a human CD1C α3 probe was used to screen several sheep thymocyte cDNA libraries. The HCD1B-like clone SCD1A25 was isolated from a foetal thymocyte library. A homologous probe comprising the α3/TM/CYT domains from this clone was derived by PCR amplification and used to identify a further three ovine clones- SCD1B-42, SCD1B-52 and SCD1T10. Three of the four clones are truncated at the 5' end, with sequences beginning towards the end of the α1 domain or the start of the α2 domain. These 5' truncation events probably reflect poor reverse transcriptase activity during library preparation. The fourth clone, SCD1B-52, represents a transcript containing a precise α3 deletion. The PCR technique was used to amplify the missing 5' ends from two of the three truncated sequences, thus generating full length coding sequence for two of the four ovine CD1's identified.

Comparison of the ovine CD1 sequences amongst themselves has shown them to be 81-96% identical at the nucleotide level and 79-90% identical at the amino acid level, suggesting that the four clones represent different gene products rather than allelic variants of CD1. The sheep sequences have also been analysed by comparison to the human, mouse and rabbit coding seqences. Perhaps unexpectedly, given the existence of five different human CD1 genes, all of the ovine CD1 sequences are most homologous to human CD1B at both the nucleotide and amino acid levels. The sheep CD1 sequences also show a high percentage sequence identity to the cottontail rabbit sequence, which is itself most similar to HCD1B.

Southern blot analysis of genomic DNA digested with a variety of enzymes and probed with the homologous α3 probe has indicated the possible existence of up to seven ovine CD1 genes. Further studies are required to determine which of these genes are expressed and to identify the genes encoding the CD1 molecules recognised by the monoclonal antibodies. The significance and implications of these results are discussed and potential further experiments suggested.

Acknowledgments.

I would first of all like to thank my Mum for her constant support and in particular for her financial support during this past year which has enabled me to complete my thesis with few financial worries.

Thanks also to Cameron for keeping me sane and for putting up with me all these years.

I am very grateful to everyone in Vet. Path. who has offered constructive scientific advice, during the course of this work. I would particularly like to thank my supervisors John Hopkins and Bernadette Dutia, as well as David Sargan, Doug Roy, Chris Woodall, Paul Hunt and Carolyn Fiskerstrand. I would also like to thank my friends both from within and without Vet. Path. for their support and encouragement during the hard times, particularly Kathy, Elaine, Penny, Vipan, Susan, Keith and Michael.

I am grateful for the support of a PhD grant from the Dept. of Agriculture, N.I. which enabled me to carry out these studies.

Part of this research was carried out in Dr. Wayne Hein's lab. at the Basel Institute for Immunology, which was founded and is supported by Hoffmann-La Roche. I am very grateful for having had the opportunity to work there and experience life in a European city. I am particularly thankful for the friendship and support which I received from Peter and Juliette Landolt-Balzarini during my stay.

Declaration

I declare that the composition of this thesis and the work presented herein are my own, except where specifically stated in the text. No part of this work has been, or is being, submitted for any other degree or qualification.

Elaine D. Ferguson

Contents.

Page Number

Acknowledgments and Declaration		i
Contents		ii
Abbreviations		vii
Chapter	**1. Literature Review**	1
1.1	Overview Of Antigen Recognition By T cells	2
1.2	T Cell Biology	2
1.2.1	$\alpha\beta$ T Cells	2
1.2.2	$\gamma\delta$ T Cells	4
1.2.3	T Cell Development	7
1.2.3.1	Positive Selection	8
1.2.3.2	Negative Selection	10
1.2.4	T Cell Activation	10
1.3	The Major Histocompatibility Complex	13
1.3.1	Classical Class I MHC Genes and Molecules	14
1.3.2	Non-classical Class I MHC Genes and Molecules	15
1.3.3	MHC Class II Genes and Molecules	17
1.4	MHC-Dependent Antigen Processing and Presentation	18
1.4.1	Processing and Presentation of Antigen in Association With Class I Molecules	18
1.4.2	Processing and Presentation of Antigen in Association With Class II Molecules	21
1.5	The CD1 Genes	22
1.5.1	Identification and Isolation of the CD1 Genes	22
1.5.2	Chromosomal Localisation of the CD1 Genes	24
1.5.3	CD1 Polymorphism	25
1.5.4	Subclasses of CD1 Genes and Evolutionary Implications	26
1.5.5	CD1 Transcription Patterns and RNA Splicing	26

1.6	The CD1 Molecules	29
1.6.1	Serological and Biochemical Characterisation of the CD1 Molecules	29
1.6.2	Cellular and Tissue Distribution of the CD1 Molecules	31
1.6.3	Molecular Associations of the CD1 Molecules	33
1.6.4	Influence of Cytokines on Expression of CD1 Molecules	34
1.6.5	Relationship Between TL and CD1	35
1.7	Function(s) of the CD1 Molecules	35
1.7.1	Class Ib Molecules as Ligands for T Cell Receptors	36
1.7.2	Peptide Presentation by Class Ib Molecules	38
1.8	Thesis Objectives	39
Chapter	**2. Materials and Methods**	**41**
2.1	Materials	41
2.2	Bacterial Cultures	41
2.2.1	Bacterial Growth Media and Plates	41
2.2.2	Bacterial Strains, Stocks and Cultures	41
2.2.3	Preparation of Competent Bacteria	42
2.2.4	Transformation Procedure	43
2.3	Preparation of Plasmid Vectors and Inserts for Cloning	43
2.3.1	Preparation of Plasmid Vector for Cloning	43
2.3.2	Preparation of DNA Fragments for Blunt-Ended Cloning	43
2.4	Ligation of DNA Fragments	44
2.5	Restriction Endonuclease Digestion of DNA Samples	44
2.6	Purification and Quantitation of DNA	44
2.7	Preparation of ^{32}P-Labelled Probe	45
2.8	Preparation of M13 Replicative Form DNA	46
2.8.1	Preparation of M13 'Phage Stock	46

2.8.2 Large-scale Preparation of M13 Replicative Form DNA 47

2.9 Screening of cDNA Bacteriophage Libraries by Colony Hybridisation 47
2.9.1 Library Plating and Nitrocellulose Lifts 47
2.9.2 Hybridisation of DNA Probes to Nitrocellulose Lifts 48
2.9.3 Secondary Screening of Bacteriophage Libraries 49
2.9.4 Preparation of Bacteriophage Lambda (λ) DNA 49

2.10 DNA Sequencing 50
2.10.1 Small-scale Preparation of Plasmid DNA for Double-stranded Sequencing 50
2.10.2 Large-scale Preparation of Plasmid DNA for Double-stranded Sequencing 50
2.10.3 Preparation of Plasmid DNA for Single-stranded Sequencing 51
2.10.4 Sequencing Reactions 52
2.10.5 Sequencing Gel Electrophoresis 52
2.10.6 Analysis of Sequencing Data 53

2.11 Southern Blot Analysis 53
2.11.1 Preparation of Genomic DNA 53
2.11.2 Southern Hybridisation 54
2.11.3 Hybridisation of Nylon Filters 54

2.12 RNA Isolation 55
2.12.1 Isolation of Poly-A^+ RNA 55
2.12.2 Isolation of Total RNA 55
2.12.3 Gel Electrophoresis of Total RNA 56

2.13 cDNA Synthesis and Tailing Reactions 56
2.13.1 cDNA Synthesis 56
2.13.2 Tailing Reactions 57

2.14 Polymerase Chain Reaction (PCR) 57

Chapter 3. Results **61**

3.1	**Library Screening and Identification of Ovine CD1 cDNA Clones**	**61**
3.1.1	Introduction	61
3.1.2	Libraries, Probes and Preparation for Use	63
3.1.3	Library Screening and Isolation of Ovine CD1 cDNA Clones	66
3.1.4	Presentation of cDNA Clone Sequencing Data	74
3.1.5	Discussion	84
3.2	**PCR Amplification of 5'UT/α1 Regions**	**95**
3.2.1	Introduction	95
3.2.2	Experimental Rationale for PCR of the 5'UT/α1 Regions	95
3.2.3	PCR Amplification of the 5'UT/α1 Regions	97
3.2.4	Presentation of 5'UT/α1 Sequence Data	100
3.2.5	Discussion	103
3.3	**Analysis of the Ovine CD1 Sequences**	**108**
3.3.1	Introduction	108
3.3.2	Overview of the Ovine CD1 Nucleotide and Amino Acid Sequence Data	112
3.3.3	Ovine CD1 Sequence Analysis	114
	A. Nucleotide Sequence Analysis	114
	B. Amino Acid Sequence Analysis	116
3.3.4	Ovine CD1 Sequences are HCD1B Homologues	119
3.3.5	Do SCD1B-42 and SCD1B-52 Represent Alleles From a Single Locus?	126
3.3.6	Relationship Between Ovine CD1 and MHC Class I Molecules	128
3.3.6.1	Interaction of Sheep CD1 α Chains with β2-microglobulin	128
3.3.6.2	Interaction of Sheep CD1 α Chains with CD8	132
3.3.6.3	Interaction of Sheep CD1 with Peptide	134
3.3.6.4	Comparison of CD1α3, Class I α3, and Class II β2 Domains	137
3.3.7	Summary and Conclusions	137
3.4	**Southern Hybridisation Analysis of Ovine CD1**	**140**
3.4.1	Introduction	140
3.4.2	Southern Blot Analysis	140

3.4.3 Results and Discussion 142

3.5 PCR Analysis of CD1 mRNA Transcription in the Intestinal Mucosa 144
3.5.1 Introduction 144
3.5.2 Experimental Rationale 145
3.5.3 PCR Protocol for Analysis of Intestinal CD1 Transcription 147
3.5.4 Discussion 149

Chapter 4. Discussion 151

4.1 The Ovine CD1 Gene Family 151
4.2 Evolutionary Relationship Between CD1 and MHC Molecules 153
4.3 What is the Function of the CD1 Molecules? 158

References 162

Appendices 197
Appendix 1: Media and Buffer Solutions 197
Appendix 2: Primer Sequences 203
A. Sequencing Primers 203
B. PCR Primers 204
Appendix 3: Genbank Accession Numbers 205

Abbreviations

ATP	adenosine triphosphate
AP	ammonium persulphate
APC	antigen presenting cell
bp	base pairs
BSA	bovine serum albumin
CD	cluster of differentiation
cDNA	DNA complementary to mRNA
CIP	calf intestinal phosphatase
CPL	compartment for peptide loading
cpm	counts per minute
CTL	cytotoxic T cell
DAG	diacylglycerol
dCTP	deoxy cytosine triphosphate
DEPC	diethylpyrocarbonate
DN	double negative
dNTP	deoxy nucleoside triphosphate
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
DP	double positive
dsDNA	doublestranded DNA
DTT	dithiothreitol
E.coli	*Escherichia coli*
EDTA	ethylenediamine tetra-acetate
ER	endoplasmic reticulum
g	gravity
g, mg, µg	gramme, milligramme, microgramme
GPI	glycophosphatidylinositol
GSP	gene specific primer
GTC	guanidinium isothiocyanate
GTE	glucose-Tris-EDTA
HPLC	high performance liquid chromarography
IAA	isoamyl alcohol
IEC	intestinal epithelial cells
Ig	immunoglobulin
Ii	invariant chain
IP_3	inositol-1,4,5-triphosphate

IPTG	isopropylthio-beta-D-galactoside
kb	kilobase
kDA	kiloDalton
LB	Luria broth
LMP	low molecular weight polypeptide
M, mM, μM	molar, millimolar, micromolar
ml, μl	millilitre, microlitre
MHC	major histocompatibility complex
MOI	multiplicity of infection
MOPS	3-(N-morpholino)-propanesulfonic acid
mRNA	messenger RNA
mya	million years ago
ODx	optical density at x nanometres
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
pfu	plaque forming units
PIP_2	phosphatidylinositol-4,5-bosphosphate
PKC	protein kinase C
PLC	phospholipase C
PTK	protein tyrosine kinase
PTPase	protein tyrosine phosphatase
RNA	ribonucleic acid
RT-PCR	reverse transcriptase PCR
SDS	sodiumdodecyl sulphate
SDW	sterile distilled water
SM	sodium/magnesium
SSC	salt/sodium/citrate
ssDNA	singlestranded DNA
ssRNA	singlestranded RNA
STE	sodium chloride-Tris-EDTA
TAE	Tris-acetate-EDTA
TAP	transporters associated with antigen processing
TBE	Tris-borate-EDTA
TCA	trichloroacetic acid
TCR	T cell receptor
TE	Tris-EDTA
TEMED	tetramethylethylenediamine

IPTG	isopropylthio-beta-D-galactoside
kb	kilobase
kDA	kiloDalton
LB	Luria broth
LMP	low molecular weight polypeptide
M, mM, µM	molar, millimolar, micromolar
ml, µl	millilitre, microlitre
MHC	major histocompatibility complex
MOI	multiplicity of infection
MOPS	3-(N-morpholino)-propanesulfonic acid
mRNA	messenger RNA
mya	million years ago
ODx	optical density at x nanometres
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
pfu	plaque forming units
PIP_2	phosphatidylinositol-4,5-bosphosphate
PKC	protein kinase C
PLC	phospholipase C
PTK	protein tyrosine kinase
PTPase	protein tyrosine phosphatase
RNA	ribonucleic acid
RT-PCR	reverse transcriptase PCR
SDS	sodiumdodecyl sulphate
SDW	sterile distilled water
SM	sodium/magnesium
SSC	salt/sodium/citrate
ssDNA	singlestranded DNA
ssRNA	singlestranded RNA
STE	sodium chloride-Tris-EDTA
TAE	Tris-acetate-EDTA
TAP	transporters associated with antigen processing
TBE	Tris-borate-EDTA
TCA	trichloroacetic acid
TCR	T cell receptor
TE	Tris-EDTA
TEMED	tetramethylethylenediamine

UT	untranslated
UV	ultraviolet
X-gal	5-bromo-4-chloro-3-indolyl-beta-D-galactoside
YT	yeast/tryptone

1. Literature Review

The vertebrate immune system has evolved to recognise and eliminate infectious material from the body. The body is protected by innate mechanisms, which are not intrinsically affected by prior exposure to the infectious agent. Higher organisms have also evolved the adaptive immune response, which is characterised by the specificity and memory of the response.

The ability to recognise foreign antigen in a specific manner is the hallmark of the adaptive immune response. There are three types of antigen recognition molecules- the immunoglobulins, the T cell receptors (TCR), and the major histocompatibility complex molecules (MHC). The genes encoding all three types of molecule are related and may have evolved from a single common precursor (Hunkapiller and Hood, 1990). Antigen recognition molecules are characterised by their diversity and heterogeneity. All three types are cell surface molecules- the immunoglobulins and TCR's act as antigen specific receptors for B and T cells respectively. In addition to cell surface expression, the immunoglobulins can be secreted as soluble proteins and act to trigger effector mechanisms. The MHC molecules are not antigen specific receptors as such, but are essential components in TCR recognition of antigen. Three classes of MHC molecules have been identified (class I, II and III), but only the class I and II gene products are directly involved in the presentation of antigen to T cells.

The CD1 genes are not encoded in the MHC gene complex but their gene products are structurally similar to the MHC class I molecules. The CD1 molecules are particularly similar to the nonclassical class I molecules in both mouse and man. This group of molecules are collectively referred to as the class Ib proteins and are characterised by limited polymorphism and restricted tissue distribution in comparison to the classical class I molecules (Stroynowski and Fischer Lindahl, 1994). The structural relatedness of CD1 molecules to MHC class I molecules, and their preferential expression on cells involved in antigen presentation to T cells, suggests that they may function in a similar manner to the class I molecules. Furthermore, it has been proposed that the CD1 molecules, like the MHC class I molecules, may be able to present antigen to T cells and in particular to $\gamma\delta$ T cells (Strominger, 1989;

Porcelli et al, 1991). This literature review therefore includes a review of T cell biology and MHC structure and function in addition to a full review of CD1 biology, including the CD1 genes and antigens, and the proposed functions of the CD1 molecules.

1.1 Overview Of Antigen Recognition By T Cells.

There are two main types of T lymphocytes- T cells which express the $\alpha\beta$ TCR and T cells which express the $\gamma\delta$ TCR. Both types of receptor are noncovalently associated with the CD3 complex. $\alpha\beta$ T lymphocytes recognise foreign antigen via the TCR/CD3 complex when the antigen is presented as a peptide in the context of a self MHC molecule. $\alpha\beta$ T cells can be subdivided into $CD4^+$ T cells and $CD8^+$ T cells. In general, $CD4^+$ T cells are referred to as helper cells (T_H) because they chiefly function to promote T and B cell proliferation and immunoglobulin synthesis, although $CD4^+$ T cells which perform cytotoxic functions have been detected. Helper T cells usually recognise antigen in association with self MHC class II molecules. The $CD8^+$ T cells are normally referred to as cytotoxic T cells (T_C) because of their ability to kill infected target cells. Again, exceptions to this "rule" have been observed i.e. $CD8^+$ T cells which perform helper functions. Cytotoxic T cells recognise antigen in association with self MHC class I molecules. The $CD4^+$ T cells can be further subdivided, particularly in mice, into TH1 and TH2 cells, depending on their cytokine secretion profiles (Mosmann and Coffman, 1989).

1.2 T Cell Biology.

1.2.1 $\alpha\beta$ T Cells.

The $\alpha\beta$ T cell receptor was identified when clone-specific monoclonal antibodies (mAbs) were raised against T cell lymphomas, hybridomas and clones (reviewed by Allison and Lanier, 1987). All of the monoclonal antibodies recognise an 80-90kD heterodimer composed of a 39-46kD α chain disulphide bonded to a 40-44kD β chain. These clone-specific mAbs are capable of affecting antigen specific activation, suggesting that they recognise the T cell antigen receptor. Co-capping experiments revealed that the clonotypically variable $\alpha\beta$TCR was associated with a group of invariant proteins called the CD3 complex (Meuer et al, 1983).

The first TCR α and β genes were identified and cloned using subtractive or differential hybridisation techniques (Yanagi et al, 1984; Hedrick et al, 1984a). The polypeptides encoded by the α and β genes have significant structural homology to the immunoglobulin genes (Hedrick et al, 1984b; Yanagi et al, 1984). The TCR α and β chains are encoded by genes constructed by rearrangement of several germline segments in a manner analogous to human and mouse immunoglobulin genes.

The variable region of the human β chain gene is assembled from three segments- variable (Vβ), diversity (Dβ) and joining (Jβ). There appear to be 75-100 Vβ gene segments, 2 Dβ regions and 13 Jβ gene segments (Toyonaga and Mak, 1987). In the mouse there are only ~20 Vβ gene segments. The human α chain gene complex consists of a large number of Vα segments (50-100) and 60-80 Jα segments, upstream from a single constant region gene (Cα). Diversity in the variable regions of the α and β chains results from germline diversity, combinatorial diversity, N region diversity (addition of extra nucleotides at junctional sites) and from multiple translational reading frames in the Dβ segment (Kronenberg et al, 1986). The potential diversity of the $\alpha\beta$ TCR repertoire has been estimated at ~10^{13} different β chain forms and 10^{12} different α chain forms (Hunkapiller and Hood, 1990).

The T cell antigen receptors are multimeric protein complexes composed of the variable α and β chains in noncovalent association with the invariant CD3 polypeptides. The CD3 complex comprises the γ, δ, ε and ζ chains (Clevers et al, 1988). The stoichiometry of the CD3 complex is thought to be two ε heterodimers, composed of either $\varepsilon\gamma$ or $\varepsilon\delta$, in association with one ζ homodimer (although $\zeta\eta$ and $\zeta\gamma$ heterodimers have also been detected) (Chan et al, 1992; Weiss, 1993). This model is based on the acidic and basic amino acid content of the transmembrane domains of each subunit, which are known to play a critical role in polypeptide chain pairing. Murine T cells are able to generate an η chain by alternative splicing of the ζ mRNA transcript (Jin et al, 1990), although there is as yet no evidence for an equivalent chain in humans. The CD3 ζ polypeptide chain is also a component of the immunoglobulin G (IgG) Fc receptor CD16, and the transmembrane segment of CD3ζ is highly

homologous to the γ subunit of the IgE Fc receptor (Terhorst and Regueiro, 1993). The α and β chains of the TCR are responsible for antigen recogniticn whilst the CD3 complex and ζ chains are involved with regulation of TCR/CD3 assembly and signal transduction. A critical role for the CD3 ζ and η gene products in signal transduction and intrathymic T cell differentiation has been established based on the results obtained from transgenic mice in which these genes were disrupted (Ohno et al, 1993; Malissen et al, 1993). Several lines of evidence have now suggested that a functional cell surface TCR is composed of two $\alpha\beta$ units in association with one CD3 complex (Chan et al, 1992; Terhorst and Regueiro, 1993).

The discovery of MHC restricted recognition of antigen by T cells (Zinkernagel and Doherty, 1974) led to the proposal of two possible explanatory mechanisms- A) that T cells have a single receptor which can interact with a combination of antigen and MHC, or B) that T cells have two separate receptors, one specific for antigen and the second for MHC. Several studies have now provided compelling evidence to show that TCR's have dual specificity for both antigen and MHC (Allison and Lanier, 1987). Conclusive evidence of the dual recognition model for both MHC class I and class II restricted T cells came when it was demonstrated that transfer of functional α and β chain genes from one T cell to another is sufficient to transfer specificity for both antigen and MHC from the first T cell to the second (Dembic et al, 1986; Saito et al, 1987).

1.2.2 $\gamma\delta$ T Cells.

During attempts to clone a murine α chain gene, a third type of T cell receptor cDNA was isolated (Saito et al, 1984). The deduced amino acid sequence of the cDNA had clear similarities to the immunoglobulin and TCR β chains in structure and sequence, and the corresponding gene was rearranged in T cell clones (reviewed by Raulet, 1989a). However, α chains were known to be glycosylated on asparagine residues, while the deduced protein sequence from this cDNA contained no consensus sites for asparagine-linked glycosylation. Soon after, authentic α chain cDNA clones were isolated and this cDNA clone was renamed "γ". In the following few years, evidence was provided for the expression of a second heteodimeric TCR, the $\gamma\delta$ TCR, which is also found in association with the CD3 complex , and

several $\gamma\delta$ T cell lines and clones were derived (Brenner et al, 1986). The δ chain gene was subsequently isolated from both man and mouse (Chien et al, 1987; Hata et al, 1987).

In man, the γ chain gene complex comprises 7-8 functional $V\gamma$ gene segments (although 14 segments, which divide into 4 subgroups, have been identified) and 5 $J\gamma$ gene segments which can rearrange with 2 $C\gamma$ gene segments (Porcelli et al, 1991). In the mouse, there are 7 $V\gamma$ gene segments which subdivide into 5 families, 4 $C\gamma$ gene segments (1 of which is a pseudogene) and 4 $J\gamma$ gene segments (Raulet, 1989). The δ chain gene complex in man contains 6 $V\delta$, 3 $D\delta$, 3 $J\delta$ and 1 $C\delta$ gene segment (Moretta et al, 1991). In the mouse, the δ chain gene locus contains 8 $V\delta$ subfamilies, several of which have only one member but some have two or more members. In addition, there are 2 $J\delta$, 2 $D\delta$ and 1 $C\delta$ gene segment. Unlike other rearrangement gene families, assembled δ genes in the human adult thymus often contain two or three D regions (Porcelli et al, 1991).

The TCR γ and δ loci contain relatively few V gene segments so that the combinatorial repertoire produced by gene recombination in the thymus was initially thought to be smaller than for the $\alpha\beta$ TCR, which utilises large pools of different gene segments (Porcelli et al, 1991). However, there is considerable potential for extensive junctional diversity in both the γ and δ gene families so that the $\gamma\delta$ TCR repertoire is actually much larger than the $\alpha\beta$ TCR repertoire. Imprecise joining can result in variable inclusion of nucleotides at all 6 DNA ends involved in these DNA rearrangements, and N regions can be added at all 3 junctions. The use of up to 3 D segments in the rearranged human δ gene also contributes to diversity. Interestingly, the $\gamma\delta$ TCR repertoire that is actually available in the periphery is restricted because cells localised at specific sites preferentially use certain V gene combinations in their receptors. In mice, $\gamma\delta$ T cells at different mucosal surfaces have distinct and limited receptor diversities. Similarly, 50-70% of human peripheral blood $\gamma\delta$ T cells express the $V\gamma9$ and $V\delta2$ gene segments (de Libero et al, 1991).

The majority of $\gamma\delta$ T cell studies have been conducted in the "$\gamma\delta$ low" species of mice and humans, and current perceptions of $\gamma\delta$ T cell diversity, development and function have been

based almost entirely on these studies. Interestingly, recent studies of γδ T cells in the sheep, a species in which γδ T cells are quantitatively more significant, have revealed that the sheep γδ T cell repertoire is more diverse than that in humans and mice due to the presence of a greater number of germline variable and constant region gene segments (Hein and Dudler, 1993). The sheep contains 28 Vδ regions belonging to 4 families and 10 Vγ regions belonging to 6 families. The ontogeny of the peripheral repertoire was found to be strictly regulated by the thymus. The authors suggest that these results, together with the prominence of γδ T cells in the sheep immune system (Hein and MacKay, 1991), argue that this T cell lineage may perform a more "elaborate function" in the sheep than in the "γδ low" species where γδ T cells may be redundant.

Despite intensive investigation, no consensus has yet emerged regarding the potential ligands for γδ TCR. To date, a heterogeneous group of γδ TCR ligands have been identified (reviewed in Haas et al, 1993) which include classical MHC proteins (Haas et al 1993), MHC-like molecules including the nonclassical MHC class Ib and CD1 molecules (reviewed in Porcelli et al, 1991), and mycobacterial and heat shock proteins (Raulet, 1989b; Young and Elliott, 1989; Kaufmann, 1990; Born et al, 1990). It has generally been assumed that γδ T cell recognition is analogous to αβ T cell recognition, although the exact nature of γδ TCR ligand recognition has yet to be established. However, the results presented in a recent paper have provided challenging evidence that γδ T cells require neither MHC class I nor class II antigen processing pathways and that peptides do not confer specificity (Schild et al, 1994). The data suggest that γδ T cells can recognise antigen in a manner akin to that of immunoglobulins and distinct from that of αβ T cells.

Much of the early work on γδ T cells focused on the possibility that these cells represent an immature stage in T cell ontogeny- γ mRNA levels were observed to be highest in thymocytes and lowest in most resting peripheral T cells (Raulet et al, 1985; Snodgrass et al, 1985). However, the discovery that γ and δ chain genes were not functionally rearranged in cloned αβ T cells raised doubts about this model (Traunecker et al 1986). Subsequent studies on γδ T cell lines and clones revealed that these cells can carry out typical T cell effector functions including cytolysis of target cells, lymphokine secretion and helper activity for immunoglobulin

production (Raulet, 1989a; Porcelli et al, 1991). For this reason it is presumed that activated γδ T cells contribute to host defence in a manner akin to αβ T cells. However, γδ T cells are distinct from αβ T cells in their localisation patterns, in their lack of demonstrable MHC restriction, and in their bias towards recognition of microbial antigens and heat shock proteins (Porcelli et al, 1991; Born et al, 1990). Despite extensive research, the role of γδ T cells in host defence against infection has yet to be defined precisely. One suggestion has been that γδ T cells may function as a first line of defence to eliminate stressed, damaged or transformed cells at epithelial sites (Janeway et al, 1988). It seems likely that the reason for our current lack of understanding regarding γδ T cell function is that most investigations have been designed on the basis that γδ T cells are just another type of αβ T cell, and further analyses will be necessary to elucidate their true function.

1.2.3 T Cell Development

The early stages of thymic development are characterised by ordered changes in expression of cell surface molecules and ordered rearrangements of TCR genes (Robey, 1994; Rothenberg, 1992). In adults, thymocyte precursors originate in the bone marrow and migrate to seed the thymus rudiment. This migration process is probably mediated by chemoattractants and homing receptors. T cell development progresses in ordered stages as thymic immigrants migrate through the cortex to the medulla, from which selected, mature T cells emerge into the periphery. The earliest immature thymocytes express Pgp-1 (CD44) and IL-2Rα (CD25), but do not express CD4 or CD8 and the TCR genes remain in the germline configuration. In human thymocytes, TCR γ and δ transcripts are detectable in the absence of the α gene transcript, which only appears later in ontogeny. It was initially suggested that T cell precursors sequentially rearrange their γ and δ loci followed by their α and β loci. However, examination of the δ gene deleted as circular DNA during α gene rearrangement showed that the δ gene was in the germline configuration, implying that these lymphocytes were already committed to the αβ lineage before TCR δ rearrangement took place (reviewed in Porcelli et al, 1991). In the αβ T cell lineage, rearrangements at the TCR β locus occur, followed by a switching off of IL-2Rα expression and gene rearrangement at the TCR α locus. The double negative ($CD4^-CD8^-$ or DN) cells progress to a double positive stage accompanied by cellular

proliferation. The double positive (DP) cells ultimately differentiate into mature single positive (SP) $CD4^+$ or $CD8^+$ thymocytes with high TCR expression and these cells are exported to the periphery.

The T cell repertoire is shaped by the processes of positive and negative selection that occur in the thymus (Rothenberg, 1992). T cells are positively selected for recognition of antigen in the context of self- MHC molecules (reviewed by von Boehmer, 1994). This process is thought to be mediated by engagement of TCR's on immature thymocytes with self-MHC expressed on thymic cortical epithelial cells (Benoist and Mathis, 1989). Only those thymocytes expressing TCR's with sufficient affinity for self-MHC are thought to be given the signal to avoid programmed cell death (Sprent and Webb, 1987). Hence, all positively selected T cells are self-MHC restricted. During this process, class I restricted T cells become $CD4^-CD8^+$, whereas class II restricted T cells become $CD4^+CD8^-$ i.e. committment to the CD4 helper lineage or to the CD8 cytotoxic lineage. Negative selection eliminates those T cells which are potentially autoreactive and is also thought to involve engagement of TCR's on immature thymocytes (reviewed by Nossal, 1994). In this case, the peptide/MHC complexes recognised by the TCR's are expressed on bone marrow derived thymic medullary stromal cells (macrophages and dendritic cells). Thymocytes expressing TCR's with too high an affinity for the peptide/self-MHC complexes are deleted by programmed cell death. The selection processes occuring in the thymus lead to the death of >95% of thymocytes. The enduring paradox for immunologists is that both positive and negative selection are mediated through the TCR but have very different outcomes.

1.2.3.1 Positive Selection.

The current model for the positive selection of thymocytes is essentially a two-step process, the first step regulating TCR gene rearrangement and the second step shaping the TCR repertoire for antigen recognition (Robey, 1994; von Boehmer, 1994). TCR β chain rearrangement has been shown to be crucial in the early phase of T cell development. Introduction of a TCR β transgene into various **rearrangement** deficient mice promoted the development of thymocytes (von Boehmer, 1994). These experiments established that a

rearranged TCR β gene is essential for the regulation of allelic exclusion of TCR β and for promoting the maturation and proliferation of thymocytes. This view was supported by the fact that >75% of TCR β rearrangements in the thymus of TCR α deficient mice were productive, indicating that a selection process had occurred (Mallick et al 1993). It now appears that early T cells display an immature pre-TCR at the cell surface which is composed of a TCR β chain disulphide linked to the developmentally regulated protein gp33, and that this heterodimer is associated with a partial CD3 complex (Groettrup et al, 1993; Groettrup and von Boehmer, 1993). It is likely that this pre- TCR complex can mediate the first stage of positive selection of DN immature thymocytes by transmitting a signal via CD3 which may require the protein tyrosine kinase, $p56^{lck}$. Consequently, further TCR β rearrangement is suppressed in the selected cells, ensuring expression of a single TCR β allele. Interestingly, the TCR α protein does not feed back to control gene rearrangement and it is possible that mature T cells may have two productive TCR α genes. Indeed, this has now been observed at both the mRNA level and at the cell surface of human T lymphocytes (Padovan et al, 1993).

The first evidence which implicated peptides in the process of positive selection was provided by studies performed to assess the effects of class I MHC mutations on T cell selection (Nikolic-Zugic and Bevan, 1990; Jacobs et al, 1990). Mutations located exclusively in the MHC groove, presumably inaccessible to the TCR, were found to influence positive selection. More recent experiments using TCR transgenic and class I deficient mice models suggest that a variety of peptides are more efficient at positive selection than a single peptide (Hogquist et al, 1994; Hogquist et al, 1993; Ashton-Rickardt et al, 1994). These studies support the idea that the endogenous peptide repertoire contains peptides which are collectively able to select an entire T cell repertoire which is able to respond to almost any pathogen (Allen, 1994).

The mechanism by which T cells with class I specific TCR's become $CD4^-CD8^+$ cytotoxic cells and T cells with class II restricted TCR's become $CD4^+$ $CD8^-$ helper T cells has been the subject of intense investigation in recent years (von Boehmer and Kisielow, 1993). Two models have been proposed- the instructive model (von Boehmer, 1986) and the

stochastic/selective model (Robey, 1994; Robey et al, 1991; Davis et al, 1993; Chan et al, 1993). The instructive model predicts that a class I MHC- specific TCR and CD8 coreceptor engaged by the same class I molecule generate a different signal from a class II MHC-engaged receptor complex, thus instructing the cell to differentiate into either a helper or cytotoxic cell precursor. In the stochastic/ selective model, lineage commitment is essentially random and by crosslinking a "matched" coreceptor to an MHC molecule i.e. CD4 with class II MHC and CD8 with class I MHC, the cell would be selected for survival. Neither model has been effectively proven or disproven and further work is necessary to elucidate the exact nature of CD4/CD8 lineage commitment.

1.2.3.2 Negative Selection.

The process of negative selection in the thymus ensures that T cells which are potentially autoreactive are eliminated, a process referred to as clonal deletion (reviewed by Nossal, 1994). In addition, other tolerance mechanisms operate in the periphery, since many self-T cell epitopes are not present in the thymus to induce clonal deletion. In the past few years, many of the investigations into the mechanisms of negative selection have been performed in superantigen or transgenic mouse model systems (Simpson, 1992; Herman et al, 1991; Kisielow et al, 1988). Induction of negative selection in the thymus is thought to be the consequence of high affinity TCR engagement of self-MHC/peptide complexes. These complexes seem to be expressed at high density on bone marrow derived stromal cells. As for positive selection, the mechanisms of thymic negative selection and peripheral tolerance have yet to be fully elucidated.

1.2.4 T Cell Activation.

The activation of antigen specific T cells is a pivotal event in the generation of many immune responses. The initiation of T cell activation involves the interaction of multiple receptor/coreceptor pairs necessary for both signal transduction and cell adhesion (reviewed in Chan et al, 1994; Terhorst and Ruguiero, 1993). The trigger for T cell activation is the engagement of the TCR by antigen/MHC complexes. Other surface molecules including CD4, CD8, CD45, CD2 and LFA-1 (CD11a/CD18) are also important in T cell activation. As a

consequence of activation, T cells undergo a programme of gene expression and *de novo* synthesis of various surface molecules, including the IL-2 receptor. Activated T cells also produce a wide range of cytokines which enable them to regulate growth, differentiation and function of cells involved in the immune response.

CD2 and CD11a/CD18 (LFA-1) are the major adhesion molecules that contribute to the binding avidity between a T lymphocyte and an APC (Terhorst and Rugueiro, 1993). APC's express the ligands for these molecules- CD58 (LFA-3) and CD54 (ICAM-1) respectively. A nonspecific adhesion interaction between a T cell and an APC allows the T cell to "screen" potential APC's. Specific interaction between the TCR and antigen/MHC leads to an intensification of adhesion. For example, activation of PKC leads to phosphorylation of the β chain of LFA-1 (CD18), initiating high affinity adhesion. Adhesion is aborted if the TCR does not specifically recognise antigen/MHC.

Strategies utilising chimeric receptors have permitted much recent progress in understanding the relationship between TCR/CD3 structure and signal transduction. The current view of the TCR complex is that it may comprise at least two autonomous signal transducing units, one which couples the ζ chains to a protein tyrosine kinase (PTK) pathway and one which couples the CD3 ε chain to a PTK pathway (Weiss, 1993; Izquierdo and Cantrell, 1992).

It has been clear for several years that the TCR is coupled to two parallel signal transduction cascades- phosphatidylinositol turnover and PTK activation. The current view is that these pathways are coupled and do not represent independent activation pathways (Terhorst and Rugueiro, 1993). The phosphatidylinositol pathway involves phosphorylation of PLC γ1 (phospho-lipase C). Activated PLC hydrolyses PIP_2 (phosphatidylinositol-4,5-bis phosphate) yielding the second messengers IP_3 (inositol-1,4,5-triphosphate) and DAG (diacylglycerol). These messengers can in turn induce the mobilisation of intracellular calcium and the activation of protein kinase C (PKC).

A second signal transduction route consists of PTK pathways. Three cytoplasmic PTK's,

$p59^{fyn}$, $p56^{lck}$ and ZAP-70, have been implicated in T cell signalling (Weiss, 1993). Fyn and lck are members of the src family while ZAP-70 is a member of the syk family (Chan et al, 1994). It has been demonstrated that the cytoplasmic domains of both the CD4 and CD8 coreceptors are tightly associated with the PTK $p56^{lck}$. The engagement of the TCR as well as CD4 or CD8 with the same antigen/MHC complex is required for optimal signal transduction- this interaction can facilitate signalling by 30-300 fold (Janeway, 1992). The role of $p59^{fyn}$ in signal transduction remains unclear, although it is known to be associated with both the CD3 ε and ζ chains. Recent studies have implicated CD45, a molecule whose surface expression is obligatory for TCR stimulation via PTK pathways, in the regulation of PTK activity (Chan et al, 1994). The cytoplasmic domain of CD45 contains two protein tyrosine phosphatase (PTPase) domains, and in model systems it can modulate the activity of CD4/CD8 associated $p56^{lck}$ by dephosphorylation of PTK's (Janeway, 1992).

The activation of T cells via the TCR/CD3 complex is, as outlined above, a very complex process. A conformational change in the $\alpha\beta$ heterodimer as a result of ligand binding is thought to induce tyrosine phosphorylation at specific motifs in CD3 and ζ chains by the src family of PTK's. Phosphorylation of these motifs recruits the ZAP-70 PTK to propagate TCR-mediated signalling events. Current studies are investigating the roles of downstream mediators of PTK and PKC action such as the ras family of GTP binding proteins and kinases such as MAP2 and c-raf (Izquiedo and Cantrell, 1992; Moodie and Woolfman, 1994).

A two signal model of lymphocyte activation was originally proposed by Bretscher and Cohn (1970), but it is only recently that putative costimulatory receptors have actually been identified. Activation of the TCR in the presence of costimulatory signals results in T cell clonal expansion and the induction of effector functions (June et al, 1994; Fraser et al, 1993). However, antigen-specific activation of the T cell in the absence of the second signal leads to a state of unresponsiveness or anergy (Jenkins, 1992). The costimulatory signal is thought to act via a distinct, as yet uncharacterised, signal transduction pathway.

The cell surface molecule CD28 has been identified as a major costimulatory molecule. It is

expressed on the majority of resting human and murine T cells (June et al, 1994). Prevention of CD28 activation using anti-CD28 antibodies during the activation of a T cell clone by antigen primed APC's results in the failure of the clone to respond to subsequent challenge. The identification of a ligand for CD28 was first reported by Linsley and coworkers when they demonstrated that antibodies to B7 could block the adhesion of B cells to CD28 transfectants (reviewed by Linsley and Ledbetter, 1993). Subsequently it was shown that cells transfected with B7 could provide the second signal to antigen or mitogen activated T cells. However, antibodies to this original B7 (now known as B7-1) did a poor job of blocking the costimulatory signal in some systems (Cohen, 1993). Furthermore, a B7-1 knockout mouse was found to have virtually no immune defects, implying that another molecule could also transmit the second signal. A second member of the B7 family, B7-2, was then identified in both humans and mice by two independent laboratories (Freeman et al, 1993a; Azuma et al, 1993). Several other studies have now provided additional evidence for the functional importance of B7-2 (Hathcock et al, 1993; Freeman et al, 1993b; Razi-wolf et al, 1993). Many of these studies have utilised a soluble form of the alternative B7 receptor CTLA-4, a molecule structurally related to CD28 which can bind to B7 with 20-fold higher affinity. B7-1 is only detected ~24-48 hours after delivery of the first signal (Lenschow et al, 1993). Since several studies have demonstrated that a T cell must receive the second signal within 24 hours after the first, it seems that B7-2, which appears virtually immediately after activation by the first signal, may be the major costimulatory molecule. Recent evidence has suggested the existence of a third member of the B7 family, tentatively named B7-3 (Boussiotis et al, 1993). Many questions regarding the costimulatory family of receptors remain to be answered. In particular, the CD28 signalling pathway requires detailed characterisation and the precise function of CTLA-4 remains to be elucidated.

1.3 The Major Histocompatibility Complex.

Self/nonself discrimination was first demonstrated in mammals by the rejection of foreign tissue grafts in mice. The genetic loci involved in this rejection process were subsequently mapped to a region known as the major histocompatibility complex (MHC). The highly polymorphic cell surface structures involved in rejection were initially characterised using

alloantibodies produced in one inbred strain of mice immunised with cells of a second strain differing only at the MHC. The MHC gene complex is located on chromosome 17 in mice (the H-2 region) and on chromosome 6 in humans (the HLA region) (Trowsdale and Owen, 1993). The human MHC genes have been mapped and the MHC molecules analysed in detail using biochemical and molecular biology techniques (Campbell and Trowsdale, 1993).

Three classes of MHC molecules have beeen identified in the MHC complexes of both mouse and man- classes I, II and III. The class III region contains a diverse collection of >20 genes which are not involved in triggering T cell activation, and are therefore not discussed here. Class I molecules are glycoproteins composed of a 44kDa alpha chain in noncovalent association with β_2-microglobulin, and are expressed on most nucleated cells. The class II molecules are heterodimers composed of a 30-34kDa alpha chain and a 26-29kDa beta chain. Class II molecules are predominantly expressed on B cells and APC's (macrophages, monocytes and dendritic cells) but may also be expressed on other cell types under particular conditions e.g. on activated human and unactivated ovine T cells. Expression of class II has additionally been detected on other cell types in certain pathological conditions as a result of induction by the cytokine IFNγ. Most class I and class II molecules are highly polymorphic structures which function to present antigen to $CD4^+$ and $CD8^+$ T cells respectively.

1.3.1 Classical Class I MHC Genes and Molecules.

The classical class I genes comprise the HLA-A, B and C loci in humans and the H-2K, D and L loci in mice. Huge advances have been made in recent years in determining the detailed structure of the class I molecules and in particular the structure of the antigen binding site. Bjorkman and colleagues (1987a; 1987b) were the first to report a crystallographic structure for a human class I molecule, HLA-A2. A putative peptide binding site identified as a deep groove formed by polymorphic parts of the $\alpha1$ and $\alpha2$ domains was observed on the top surface of the molecule. The sides of the groove were formed by two alpha helices and the base was an eight stranded beta pleated sheet. Polymorphic residues were found to be concentrated at this binding cleft. Extra electron density within the groove was proposed to represent the image of a mixture of peptides bound by the HLA-A2 molecule. The refined

structure of HLA-A2 and the crystal structure of a second class I molecule, HLA-A68, provided further details of the peptide groove structure (Saper et al, 1991; Garrett et al, 1989). Six pockets or subsites were identified (A-F) within the groove, suggesting that polymorphism, by creating and altering such sites, can influence the range of peptides bound by a particular class I allele. The structure of a third class I molecule, HLA-B27, revealed a clearer picture of the peptides in the groove, indicating that peptides bind in an extended conformation with tight binding at either end of the groove between the peptide and conserved MHC residues (Madden et al, 1991; Madden et al, 1992). Solution of the crystal structure of the mouse class I molecule H-2K^b in association with two different bound peptides (H-2K^b+ an octamer from vesicular stomatitis virus, and H-2K^b+ a nonamer from Sendai virus) provided further details of the interaction between class I molecules and peptides (Zhang et al, 1992; Matsumara et al, 1992; Freemont et al, 1992). It seems that peptides bound to class I molecules are restricted in length to eight or nine residues. The amino and carboxy termini are held tightly within conserved pockets at either end of the cleft, while the extra length of a nine residue peptide can be **accommodated** by protrusion of residues in the middle of the peptide out of the groove (Parham, 1992).

1.3.2 Non-classical Class I MHC Genes And Molecules.

In addition to the polymorphic classical class I genes, the human and mouse MHC regions also contain a substantial number of nonclassical class I genes. These nonclassical genes are characterised by limited polymorphism, restricted tissue distribution and low cell surface expression (Stroynowski, 1990). Murine nonclassical class I molecules are encoded in the Q, T and M regions (formerly the Qa, TLa and Hmt regions) and human nonclassical class I molecules are encoded in the HLA-E, F, G, H, I and J regions (Shawar et al, 1994). Genes homologous to the murine Q and TL region genes have additionally been detected in the rat (Kirisits et al, 1994; Rothermel et al, 1993).

The H-2Q region of the mouse MHC is composed of at least 10 genes, of which the Q4 and Qa-2 genes are expressed in a wide variety of tissues (reviewed by Morse et al, 1990). The Q2, Q5 and Q10 gene products have a more tissue specific expression pattern. The Q4, Qa-

2 and Q10 gene products can be detected as secreted molecules. It has recently been reported that two forms of soluble Qa-2 exist- one form corresponds to the translation product of a truncated transcript, while the second form is derived from membrane-bound molecules (Tabaczewski et al, 1994). The membrane-bound form of Qa-2 is attached to the cell surface via a GPI (glycophosphatidylinositol) linkage and has been implicated in transmembrane signalling (Stroynowski, 1990).

The H-2T region of the mouse MHC contains at least 20 genes including the T23 gene which encodes the inappropriately named Qa-1 molecule (reviewed by Chorney et al, 1990). T region gene products are predominantly expressed on normal thymocytes and thymic leukemia cells. Several studies have also reported the expression of T region gene products in the intestinal epithelium where they have been suggested to function as presentational elements for $\gamma\delta$ T cells (Wu et al, 1991; Herschberg et al, 1990; Eghtestady et al, 1992). Furthermore, cell surface expression of a transfected Qa-1^b molecule was stabilised following heatshock and by the addition of a tryptic mycobacterial digest (Imani and Soloski, 1991), suggesting a role in peptide presentation.

The most thoroughly studied nonclassical class I genes are the M region genes of the mouse MHC (reviewed by Fischer Lindahl et al, 1991). This region of the mouse MHC is presently known to contain 8 genes (Wang et al, 1993). H-2M3 is the best understood example of a nonclassical class I presentational element. The minor histocompatibility antigen Mta is a complex of MTF, HMT and β_2-microglobulin. MTF is a polymorphic N-formylated hydrophobic peptide derived from the amino terminus of the mitochondrially encoded ND1 protein, and H-2M3 is a nonclassical class I gene. A peptide must fulfil three requirements for binding to HMT- 1) have an N-formyl moiety at the amino terminus, 2) have a space filling R group at position 1, and 3) have the correct sterioisomer of the first amino acid (Vyas et al, 1992; Shawar et al, 1990). Since bacterial as well as mitochondrial protein synthesis is initiated by an N-formyl methionine residue, it was suggested that HMT may function as a restriction element for peptides of bacterial origin (Shawar et al, 1991; Fischer Lindahl, 1991). It has subsequently been demonstrated that HMT can indeed present bacterial peptides to CTL during *Listeria*

monocytogenes infection (Pamer et al, 1992; Kurlander et al, 1992). Cell surface expression of HMT has recently been shown to be dependent on the availability of appropriate endogenous peptides (Vyas et al, 1994). Wang and colleagues have proposed that the HMT molecule may be considered as a neoclassical class I molecule, since the developmental expression and pattern of cytokine induction parallel that observed for classical class I molecules (Wang and Fischer Lindahl, 1993a).

The human HLA-E, F and G nonclassical class I regions are much less well understood than their murine counterparts. HLA-F transcripts have been detected in large quantities in foetal liver and at much lower levels in other tissues (Houlihan et al, 1992). The HLA-E gene is transcribed at low levels in most adult tissues and in the placenta (Shawar et al, 1994). HLA-G transcripts have been detected in the trophoblasts of the placenta, and more recently, alternative splicing products have been observed in adult peripheral lymphocytes and in keratinocytes (Kirszenbaum et al, 1994; Ulbrecht et al, 1994; Houlihan et al, 1992).

The putative functions of nonclassical class I molecules have been extensively debated down the years (Lawlor et al, 1990; Klein et al, 1983). Klein and colleagues have long held the view that the nonclassical class I genes constitute an evolutionary junkyard (Klein et al, 1991; Ram and Tyle, 1991). They consider that these genes have no functional significance and that they represent relics of genes once used in antigen presentation or are the products of unsuccessful attempts at producing new genes. It has been proposed that these genes may provide a source of donor sequences for the generation of sequence variation in classical class I genes (Lawlor et al, 1990). Others have suggested that the nonclassical genes may function in a manner analogous to the classical class I molecules (Strominger, 1989; Srivastava and Lambert, 1991). Various studies in recent years have now provided evidence for a presentational role for these molecules, as discussed more fully in conjunction with CD1 function in section 1.7.

1.3.3 MHC Class II Genes And Molecules.

The class II genes comprise the HLA-D region in humans (three loci-DP, DQ and DR) and H2-

IA and IE in mice. The three dimensional structure of the class II molecule HLA-DR1 has recently been determined and is similar to that of class I molecules, as previously predicted (Brown et al, 1993; Brown et al, 1988). Peptides are bound in an extended conformation with the termini projecting from both ends of an open ended antigen binding groove. One prominent peptide side chain at ~position 3 of a 15-mer identifies a critical binding pocket, and other pockets are indicated by patches of polymorphic residues in the binding site. The major difference in comparison to class I structures is that DR1 crystallises as a dimer of the $\alpha\beta$ heterodimer i.e. a dimer of dimers. Site-directed mutagenesis would be required to rule out the possibility that this dimerisation is a crystallisation artefact. It has been suggested that the dimer of dimers may influence T cell signalling by affecting CD4 binding and TCR crosslinking (Ploegh and Benaroch, 1993; Brown et al, 1993).

1.4 MHC-Dependent Antigen Processing And Presentation.

The MHC class I and II molecules serve as "peptide binding and display proteins" (reviewed by Germain, 1994). The two classes of MHC molecules are specialised for the binding and presentation of peptides in distinct intracellular locations. In general, MHC class I molecules bind to peptides in the ER (endoplasmic reticulum) which have been derived from proteins actively synthesised in the cell or entering the cytosol. In contrast, MHC class II molecules primarily interact with peptides in the endocytic pathway derived from exogenous protein sources.

1.4.1 Processing And Presentation Of Antigen In Association With Class I Molecules.

Much of the current knowledge of class I associated processing pathways has been derived from work with mutant cell lines which have defects in class I heavy chain/β_2m assembly. For example the mutant cell line RMA/S expresses 95% less class I molecules at the cell surface than the normal cell line RMA. Addition of exogenous peptides to RMA/S cells restored cell surface class I expression by stabilising the heavy chain/β_2m interaction (Monaco, 1992; Townsend et al, 1989). In the absence of peptide, the class I heavy chain does not fold in the correct conformation and is consequently unstable. The data imply that the defect in the

mutant cells arises from a lack of peptides in the ER which are necessary for effective class I assembly and cell surface expression. It was subsequently discovered that the defect in the RMA/S cell line, and in a second mutant cell line 721.134, could be corrected by transfection of cDNA encoding the TAP1 or TAP2 proteins (Spies and Demars, 1991; Attaya et al, 1992), providing strong evidence that TAP mediates peptide translocation into the ER. Transporter associated proteins (TAP) are members of the ABC (ATP binding cassette) family whose genes map to the class II region of the MHC (Deverson et al, 1990; Trowsdale et al, 1990; Spies et al, 1990). The TAP1 and TAP2 proteins are thought to form a heterodimer in the membrane of the ER (Kleijmeer et al, 1992). Polymorphisms in the TAP genes result in an alteration of the spectrum of class I bound peptides (Powis et al, 1992a; Powis et al, 1992b), providing further indirect evidence for the role of TAP in peptide translocation. It has been suggested that TAP may function as a molecular ruler, selecting those peptides of an appropriate length and with a suitable carboxy terminal residue for binding to a class I molecule (Germain, 1994; Momburg et al, 1994; Neefjes et al, 1993). However it has yet to be directly demonstrated that TAP1 and TAP2 constitute an active peptide transporter of cytosolic peptides.

Cytosolic peptides can additionally enter the ER in a TAP independent manner via the normal SRP(signal recognition particle) dependent pathway (Forquet et al, 1993) and signal sequences of ER- targeted proteins are found to be presented by HLA-A2 molecules (van Bleek and Nathenson, 1993). However, the contribution of this TAP-independent pathway to antigen presentation during an immune response is not yet clear.

Two additional genes encoded in the MHC class II region adjacent to the TAP genes have been implicated in the cytosolic degradation of proteins to peptides. The LMP2 and LMP7 proteins, discovered by Monaco and McDevitt (1982), encode two subunits of a large cytoplasmic structure known as the LMP (low molecular weight polypeptide) complex. This complex is structurally and biochemically related to the proteasome, a multicatalytic proteinase complex which functions to degrade ubiquinylated proteins in the cytosol (Brown et al, 1991). However, two reports showed that defective class I expression in a mutant cell line lacking the

genes for both TAP and LMP2/7, could be restored by transfection of the TAP genes alone, suggesting that the LMP genes are not crucial for the class I processing pathway. Several studies have now suggested that the LMP genes can influence the cleavage specificity of the proteasome and may favour the generation of peptides with appropriate carboxy termini for loading to class I molecules (Driscoll et al, 1993; Howard and Seelig, 1993).

The efficient assembly and transport of class I molecules to the cell surface is critically dependent on the presence of suitable peptides. Peptide-free class I heavy chains appear to be associated with the ER protein calnexin which prevents movement of class I through the ER (Germain, 1994). Dissociation from calnexin and formation of stable heavy chain/ β_2m/peptide complexes allows class I transport to the cell surface to proceed.

The typical view of class I processing and presentation is that proteasomes degrade cytosolic proteins to 8 or 9mer peptides which are then transported into the ER by ATP- dependent TAP transporters prior to association with class I molecules. However, processed peptide ligands for class I molecules cannot be detected in the cytosol. Furthermore, in order to be selected by class I molecules, the entire pool of peptides would need to be transported into the ER. Rammensee and colleagues have proposed an alternative model of class I processing which assumes an instructive role for class I molecules (Rammensee et al, 1993a; Rammensee et al, 1993b; Howard and Seelig, 1993). They suggest that proteins are first degraded into precursor peptides with appropriate carboxy termini for class I binding. These peptides would then be transported into the ER and interact with low affinity via the carboxy terminal residue to class I molecules. The amino terminus would then be trimmed allowing the peptide to snap into the groove and bind with high affinity. This model would require the transport of only a subset of available peptides and would fit with the current evidence for selectivity of the peptide transporters.

Huge advances have been made in recent years in the analysis of peptides naturally presented by class I molecules. Peptides can now be eluted from purified class I molecules, separated by HPLC and their sequences analysed by Edmann degradation and/or mass spectrometry. These studies have revealed that each class I allelic molecule has its own

peptide specificity or peptide motif (van Bleek and Nathenson, 1993; Hunter et al, 1992; Falk et al, 1991). Peptides bound to class I molecules are restricted to 8 or 9 residues in length, and two of the 8 or 9 positions are typically occupied by a single amino acid or anchor residue. Most of these studies have performed pool sequencing of eluted peptides, although sequencing of individual peptides is now possible (Joyce et al, 1994).

1.4.2 Processing And Presentation Of Antigen In Association With Class II Molecules.

MHC class II molecules are $\alpha\beta$ heterodimers which assemble in the ER and become complexed with a third protein, the nonpolymorphic invariant chain (Ii). The nine chain complex formed is composed of three class II dimers and three Ii chains (Lamb and Cresswell, 1992). Various functions have been ascribed to the Ii chain and it is clear that it plays a central role in class II assembly and transport. The Ii chain has been shown to facilitate class II folding and transport through the ER, and to interfere with peptide/class Ii association (Roche and Cresswell, 1990; Sant and Miller, 1994; Cresswell, 1994) suggesting that it may act as a surrogate peptide to prevent peptide interaction prior to reaching the acidic peptide loading compartment. Recent studies have revealed that a particular region of the Ii chain, the CLIP (class II associated invariant chain peptide) region, interacts with the class II binding site to prevent class II/peptide association during early biosynthesis (Riberdy et al, 1992). The Ii chain is also thought to contain, in the cytoplasmic domain, an endosomal localisation signal which directs and/or retains class II molecules within the endocytic pathway (Loss and Sant, 1993; Cresswell, 1994). Two groups have recently developed Ii negative transgenic mice (Viville et al, 1993; Bikoff et al, 1993) which exhibit dramatic defects in class II biosynthesis and function. These mice have decreased class II cell surface expression, altered class II transport and inefficient selection of CD4^{+} T cells. The results confirm the importance of the Ii chain in class II assembly and transport.

The precise location and nature of the acidic peptide loading compartment for class II molecules has, until very recently, been ill-defined. The prevailing view has been that class II/Ii complexes meet unprocessed antigen in an endosomal compartment and together proceed

along a proteolytic pathway so that unoccupied class II molecules are produced simultaneously with antigenic peptides. However, several papers published in May this year have reported the identification of a unique subcellular compartment, the CPL (compartment for peptide loading), in which class II peptide loading occurs (Amigorena et al, 1994; Tulp et al, 1994; West et al, 1994; ~~Kozono et al, 1994~~). The CPL have no detectable endosomal or lysosomal markers and have distinguishable subcellular fractionation behaviour (Schmid and Jackson, 1994). Most class II molecules in the CPL are not associated with the Ii chain. It has been suggested that CPL undergo a maturation process whereby Ii is degraded, class II dimers assume an intermediate conformation and peptide loading subsequently occurs.

Sequencing of endogenous peptides eluted from MHC class II molecules has indicated that, in contrast to class I molecules, the peptides bound are much more variable in length, typically 12-24 residues long (Rudensky et al, 1991; Hunt et al, 1992). The majority of peptides are derived from transmembrane or secreted proteins present in large quantities in the endosomal pathway (Rudensky and Janeway, 1993). Class II peptides also have variable amino and carboxy termini which, together with the length variability, indicates that specific interactions between the peptide and class II molecule occur in the central part of the peptide rather than at the termini as for class I molecules (Forquet et al, 1993).

1.5 The CD1 Genes.

The CD1 molecules are a family of β_2m-associated glycoproteins with broad structural homology and weaker sequence homology to the MHC class I antigens. The CD1 molecules are characterised by limited polymorphism, low surface expression and a restricted pattern of tissue distribution in comparison to the classical MHC class I gene products. The genes encoding the CD1 molecules have now been characterised in a variety of species.

1.5.1 Identification And Isolation Of The CD1 Genes.

The first human CD1 cDNA clone, encoding the CD1a molecule, was isolated from the cell line NH17, a derivative of MOLT4 which expresses high levels of HTA-1, the human thymocyte antigen recognised by the mAb NA1-34 (Calabi and Milstein, 1986). Two CD1a-encoding

clones were identified- FCB6 and FCB1. There are three single base differences, which are probably cloning artefacts, between the two clones, and the FCB1 clone extends ~180 nucleotides further in the 5' direction. The FCB1 clone sequence begins near the end of the α1 domain. A Southern blot of human MOLT4 DNA digested with EcoR I and probed with the insert from FCB6 identified five hybridising bands (Martin et al, 1986). These were found to correspond in size to the EcoR I fragments of five different clones isolated from a genomic library (R4B3, R1L5, R7L4, R3G1, and R2G4), and represent the five different human genes (HCD1A, B, C, D, and E). The α3 domains of the genomic clones were sequenced and the FCB6 cDNA clone was matched with the R4B3 genomic clone. Three of the genomic clones have been found to encode the serologically defined CD1a, b, and c antigens, by analysing a series of transfectants with a panel of monoclonal antibodies (mAbs) (Martin et al, 1987). cDNA clones corresponding to the CD1A, B and C genes were identified and isolated using anti-CD1 mAbs to screen an HPB-ALL cDNA library expressed in COS cells (Aruffo and Seed, 1989). A cDNA clone corresponding to the human CD1D gene was subsequently isolated from a thymus library using a mixture of full-length cDNA CD1A, B and C probes (Balk et al, 1989a). The same probes, and gene specific probes corresponding to the 5'UT, α1 and 3'UT regions, were used to identify an equivalent genomic clone. The full-length DNA sequences from the genomic clones R2G4(CD1E) and R3G1 (CD1D), isolated by Martin and colleagues, were also reported at this time (Calabi et al, 1989a). All five human sequences are highly homologous in the region encoding the membrane-proximal α3 domain (80-94% identity at the nucleotide level and 71-88% identity at the amino acid level), which is likely to be the site of interaction with β2m (Calabi et al, 1991). Comparison of the 5'UT, α1 and α2 domains indicates that CD1D is the most divergent member of the CD1 gene family (Balk et al, 1989a). Two mouse CD1 genes have been identified, both corresponding to the human CD1D gene (~76% nucleotide identity and ~65% amino acid identity; Bradbury et al, 1988). The MCD1.1 gene was isolated from a genomic library screened with the insert from the human CD1A clone FCB6 (Bradbury et al, 1988). The second gene, MCD1.2, was identified by screening a genomic library with an α3 region subclone of the MCD1.1 gene. cDNA clones corresponding to both murine genes have also been isolated using a probe corresponding to the α2 and α3 domains of the human CD1D gene (Balk et al, 1991a). There are only 19 amino acid differences in the extracellular domains of the derived murine CD1 polypeptide sequences. In

the nucleotide sequences, the homology between the two genes extends to the 5' UT region and to the first ~120 nucleotides of the 3'UT region, but the sequences in the 3'UT region then diverge (Balk et al, 1991a).

The sequence of another rodent CD1 gene, rat CD1, has recently been reported (Ichimiya et al, 1994). The cDNA clone encoding rat CD1 was isolated from a thymic lymphoma cDNA library using a rat $\alpha3$ region probe. This probe was generated by PCR amplification using primers based on conserved sequences in other CD1 genes. Southern blot analysis using the $\alpha3$ region probe indicates the existence of a single rat CD1 gene. The rat CD1 sequence is highly homologous to both murine sequences in the 5'UT region and in the coding domains. Sequence divergence in the 3'UT region indicates that rat CD1 is equivalent to the murine CD1.1 gene (80% identity) and not to the CD1.2 gene (52% identity).

Two rabbit CD1 genes have been identified, although Southern blot analysis indicates the existence of up to 8 genes (Calabi et al, 1989b). The cottontail rabbit genomic CD1 clone encodes the $\alpha1$, $\alpha2$ and $\alpha3$ domains, and is most related to the human and murine CD1D genes. The domestic rabbit genomic CD1 clone encodes the 5'UT and $\alpha1$ domains only and is most similar to human CD1B.

1.5.2 Chromosomal Localisation Of The CD1 Genes.

The chromosomal localisation of the CD1 genes has been determined in both humans and mice. Analysis of the pattern of segregation of CD1 and MHC genes in mouse x human somatic hybrids by Southern blot analysis provided the first evidence that the CD1 genes do not map to the MHC region (Calbi and Milstein, 1989b). *In situ* hybridisation using gene-specific probes and a sensitive fluorescent detection system was subsequently used to assign all five human genes to chromosome 1q22-23, confirming that they are not MHC linked (Albertson et al, 1988). The human genes constitute a tight linkage group spanning ~190kb which has been mapped using 14 overlapping cosmid clones (Yu and Milstein, 1989). The linkage of the genes was established using pulse field gel electrophoresis, cosmid cloning

and walking techniques. The genes are arranged in the order CD1D, A, C, B, E. It is interesting that the classical CD1 genes, encoding the serologically defined CD1a, b, and c antigens, are flanked by the most divergent genes in the family, and that CD1D is 2-3 times further away from CD1A in comparison to other intergenic distances (20-30kb). All of the human CD1 genes are in the same transcriptional orientation with the exception of CD1B.

The CD1 genes of the mouse are also non MHC-linked and have been mapped to chromosome 3 using Southern blot analysis of hamster/mouse somatic cell hybrids (Bradbury et al, 1991). The two mouse genes are arranged in a tail to tail orientation and are <10kb apart (Bradbury, PhD Thesis, 1987). The recently isolated rat homologue of CD1 is also unlikely to map to the MHC region, since no correlation was observed between MHC haplotype and CD1 RFLP analysis (Ichimiya et al, 1994).

1.5.3 CD1 Polymorphism.

The extent of genetic polymorphism in the CD1 genes is extremely limited. Southern blot analysis of EcoR I digested genomic DNA with an $\alpha3$ region probe revealed the existence of 5 genes, and the same five EcoR I fragments were detected in 15 individuals (Calabi and Milstein, 1986; Martin et al, 1986). A number of CD1 genomic and cDNA clones derived from different individuals have now been sequenced- three for CD1A, and two for each of CD1B, C and D (Calabi and Milstein, 1986; Martin et al, 1986; Balk et al, 1989; Aruffo and Seed, 1989). Only minor differences, mainly single base substitutions, have been detected within the coding regions of these sequences and it is likely that these represent cloning and/or sequencing artefacts.

In the mouse, Southern blot analysis of 11 inbred laboratory mice strains using an $\alpha3$ probe detected only one RFLP (restriction fragment length polymorphism) (Bradbury, PhD Thesis, 1987). A more sensitive analysis was conducted utilising nuclease protection assays and exon specific probes (Bradbury et al, 1988). This analysis was expected to detect most multiple mismatches and some but not all single base differences within the probe length. Two independent variants of the $\alpha3$ domain were detected, probably involving single base

changes.

The extent of CD1 polymorphism in the rat was similarly assessed by Southern blot analysis in 12 strains of laboratory rat using the 5'UT/α1/α2/α3 and TM/CYT/3'UT probes (Ichimiya et al, 1994). Three allelic variants of the rat CD1 gene were detected using the TM/CYT/3'UT probe. However, the nature of this probe did not enable variation in the TM/CYT domain to be distinguished from variation in the 3'UT domain.

1.5.4 Subclasses of CD1 Genes And Evolutionary Implications.

Comparison of the 5'UT, α1 and α2 domain sequences suggests that the CD1 genes can be divided into two distinct classes- the classical CD1 genes (human CD1A, B and C, and the domestic rabbit CD1B) and the human, mouse and cottontail rabbit CD1D genes (Calbi et al, 1991). The HCD1D gene is significantly more related to the murine CD1 genes than to any of the other human genes (Calabi et al, 1989a). It was suggested that these results indicate the generation of the CD1 gene family prior to primate/rodent divergence ~80 mya and that the classical CD1 genes (CD1A, B and C) have been deleted in the rodent lineage. The separation into CD1D-like and non CD1D-like genes suggests a functional dichotomy within the CD1 gene family. The CD1E gene is in an intermediate position between the classical genes and the CD1D-like genes. The two classes of genes can be distinguished by Southern blot analysis using α1 or α2 exon probes. A subsequent analysis of CD1 gene sequences and the construction of phylogenetic trees for evolutionary analysis was performed by Hughes (1991). The results of this study support the hypothesis of Calabi and colleagues (1989a) that the CD1 genes fall into different classes which arose by gene duplication prior to the divergence of the mammalian orders. However, in contrast to the Calabi results, HCD1B was found to represent a separate lineage which diverged prior to the HCD1A/C grouping. The Hughes analysis suggests that the CD1 genes diverged at around the time of the bird/mammal separation, ~250-300 mya.

1.5.5 CD1 Transcription Patterns And RNA Splicing.

Investigations of CD1 mRNA synthesis and CD1 RNA splicing complexity have been

performed in various tissues and in several animal species. Northern blot analysis was performed on RNA from the MOLT4 cell line and from the daughter cell line NH17 (which expresses CD1a at a high level) (Calabi and Milstein, 1986). The 3'UT CD1A-specific probe detected two main bands of ~2.1 and 1.4-1.6kb, although the same probe detected only one hybridising band by Southern blot. These results indicate the existence of multiple CD1A mRNA transcripts derived from a single gene. Northern blot analysis of total RNA hybridised with CD1A, B and C cDNA probes revealed the presence of CD1 transcripts in thymocytes and several T cell leukemias but not in mature T or B cells, nor in myeloid-derived tumour lines (Aruffo and Seed, 1989). Similar hybridisation patterns and intensities were obtained with all three probes, indicating that CD1A, B and C are coordinately expressed in different cell lines. The CD1 transcripts detected ranged in size between 1.2 and 3.0kb suggesting the existence of RNA splicing complexity. The CD1D gene is not transcribed to the same extent as the classical CD1 genes at the mRNA level (Balk et al, 1989). Northern blot analysis of thymus tissue using full length CD1D and CD1A probes revealed the presence of both types of transcript but with significantly less signal for the CD1D probe, suggesting that CD1D is a relatively low abundance thymic transcript. In addition, low or undetectable levels of CD1D gene transcription was observed in several T lymphoblastoid cell lines and no message in a series of B cell or myeloid lines.

The existence of a complex array of alternatively spliced human CD1 transcripts has also been suggested by both cDNA sequence analysis and nuclease protection assays (Calabi et al, 1989c). Cryptic splice sites have been detected in the leader-α1 intron of HCD1A, in the leader exon and leader-α1 intron of HCD1E, and within the α3 exons of HCD1B and HCD1E (Calabi et al, 1991). Most of these splicings cause frameshifts leading to chain termination or probable destruction of β2m binding and membrane expression. Human CD1 transcription has most recently been analysed in mouse myeloma transfectants and thymocytes (Woolfson and Milstein, 1994). RT-PCR analysis was performed on the clone 10B3 derived from a mouse myeloma cell line transfected with a partially spliced CD1A cDNA. Four PCR products were detected using primers derived from the α3 and 3'UT regions. One PCR band was found to correspond to an unspliced transcript with a cryptic stop codon in the intron between the α3

and TM/CYT domains. Protein sequencing of CD1a purified from the transfectant supernatant indicated that the unspliced transcript encodes a secretory isoform of CD1a. A second PCR band was found to correspond to an alternatively spliced transcript in which a cryptic splice acceptor site within the transmembrane/cytoplasmic exon is utilised in preference to the site at the beginning of the exon. The product of this transcript was found to be an intracellular component. A third band represented the transcript encoding the correctly spliced membrane-bound form of CD1a, and the fourth band was a heteroduplex derived from two of the other bands. A similar analysis of all five HCD1 genes was performed in thymus tissue. A single major band, corresponding to the correctly spliced membrane-bound form of CD1, was detected following amplification of CD1A and CD1B cDNA. The single CD1A transcript detected in thymus tissue contrasts sharply with the complex splicing pattern observed in the CD1A transfectant. CD1D amplification gave one major band and two additional minor bands. In contrast, CD1C and CD1E gave complex splicing patterns. The CD1C pattern was comparable to that observed in the CD1A transfectant, with unspliced, alternatively spliced and correctly spliced transcripts being detected. Human CD1 gene transcription is thus characterised by considerable mRNA splicing complexity which may be tissue specific.

Analysis of murine cDNA clones derived from total thymus RNA has revealed that the majority represent partially spliced RNA transcripts (Bradbury, PhD Thesis, 1987; Bradbury et al, 1990). RNase mapping experiments with probes containing each of the domains of mouse CD1D with a section of adjacent intron, showed that for mouse CD1 there was a significant degree of unspliced message (20-50% of total) for the $\alpha1$, $\alpha2$ and $\alpha3$ regions in the thymus, a small degree (1-5%) in the spleen, and normal splicing in the liver (Bradbury, PhD Thesis, 1987). CD1 message was detected by RT-PCR in both mouse and rat IEC (intestinal epithelial cells) at an equivalent low intensity (Ichimiya et al, 1994), in contrast to previous results using mAbs which revealed prominent staining in murine intestinal epithelium (Bleicher et al, 1990). However, the RT-PCR results are supported by *in situ* hybridisation analysis of murine intestinal epithelium which demonstrated that only Paneth cells at the bottom of crypts of Lieberkuhn express CD1 message and no signal was detected in other IEC (Lacosse and Martin, 1992).

CD1 gene transcription in the rat was analysed by Northern blot analysis using a TM/CYT/3'UT probe (Ichimiya et al, 1994). CD1 mRNA transcription was detected in nonlymphoid organs (liver, heart, kidney and lungs) as well as in the lymphoid organs of the spleen and, to a lesser extent, the thymus. Brain and IEC tissue were negative in this analysis. CD1 transcription was, however, detected in the rat intestine, and also in two thymic epithelium cell lines, using the more sensitive RT-PCR technique. The level of CD1 transcription in the IEC was comparable to that observed in the murine intestine. The level of CD1 transcription was much higher in the rat thymus than in the murine thymus- rat CD1 was readily detectable by Northern blot compared to CD1 in the murine thymus which was only detectable by S1 nuclease mapping (Bradbury et al, 1990). Northern blot analysis with the TM/CYT/3'UT probe gave no indication of splicing complexity for rat CD1 RNA in contrast to the variation in splicing of the human CD1 genes in this region.

1.6 The CD1 Molecules.

The first human differentiation antigen to be defined by a monoclonal antibody (mAb) was the CD1a molecule, then called HTA-1 (human thymocyte antigen 1)which was recognised by the mAbs NA1/34 (McMichael et al, 1979) and T6 (Reinherz et al, 1980). Serological and biochemical studies on the antigens recognised by these and several other mAbs allowed the CD1 mAbs to be clustered into three distinct groups (reviewed by Calabi et al, 1991). The mAbs recognise three different CD1 molecules, CD1a, b and c, characterised by their molecular weights, epitopic sites and tissue distributions. In the past few years, a fourth CD1 molecule, CD1d, has been characterised in humans and mice using specific mAbs (Bilsland and Milstein, 1991; Bleicher et al, 1990; Balk et al, 1994). The CD1 molecules have now been described in a variety of species including humans, mice, rats, sheep, cattle and pigs.

1.6.1 Serological And Biochemical Characterisation Of The CD1 Molecules.

Human CD1a, b and c molecules have been characterised using a panel of mAbs and various serological and biochemical techniques, including SDS-PAGE, sequential immunoprecipitations, 2D gels and analysis of IEF patterns (Terhorst et al, 1980; van Agthoven and Terhorst, 1982; van de Rijn et al, 1983; Lerch et al, 1983; Cotner et al, 1982;

Amiot et al, 1986; Knowles and Bodmer, 1982; Small et al, 1987; Calabi et al, 1991). The CD1 molecules consist of a glycosylated α chain of variable molecular weight in noncovalent association with β_2m. The CD1a, b and c α chains have molecular weights of 49kD, 45kD and 43kD respectively. The more recently characterised human CD1d molecule is also a β_2m associated protein and has an α chain of 49kD molecular weight (Bilsland and Milstein, 1991; Blumberg et al, 1991). CD1a and d α chains each have 4 potential N-linked glycosylation sites, while CD1b and c each have 3 potential N-linked glycosylation sites. The differences in the molecular weights of the α chains are due to variation in the extent of glycosylation. The panel of human anti-CD1 Abs have been used to define 7 epitopes on the three CD1 molecules CD1a, b and c- 4 epitopes are restricted to CD1a, 2 epitopes to CD1b and 1 epitope to CD1c (Amiot et al, 1987). Recent investigation into human CD1d expression in the intestinal epithelium have demonstrated the expression of a non-β_2m associated 37kD protein with no N-linked carbohydrate modifications (Balk et al, 1994). The significance of this CD1 isoform in the intestinal epithelium is unclear. A protein product of the human CD1E gene has not yet been defined, although genomic sequence analysis suggests that this gene is functional (Calabi et al, 1989).

A cluster of mAbs have similarly been used to define and characterise the murine CD1d molecules (Bleicher et al, 1990; Mosser et al, 1991). A rabbit antiserum raised against the murine CD1.2 fusion protein recognised a set of four glycoproteins (49-55kD) which contain a common core protein corresponding to the size predicted by sequence analysis (36kD, Bradbury et al, 1988) (Mosser et al, 1991). The multiple molecular weight sizes of the α chain appear to reflect varying degrees of glycosylation which may be tissue specific. The murine CD1.1 cDNA was transfected into mouse fibroblasts and used to raise the mAbs 3C11 and 1H1 (Bleicher et al, 1990). Both mAbs immunoprecipitated a 48kD α chain in association with β_2m. The molecular weight of the α chain is compatible with the accommodation of 5 potential N-linked glycosylation sites, as predicted by sequence analysis. The existence of a β_2m-associated CD1 molecule in the rat has recently been demonstrated using a rabbit anti-rat Ab raised against recombinant rat CD1 protein (Ichimiya et al, 1994). The predicted size of the polypeptide backbone is similar in both rat and mouse, although the extent of glycosylation at the 5 N-linked glycosylation sites may differ, since the rat α chain has a molecular weight of

45kD compared to 48kD in the mouse.

CD1 molecules have also been defined by mAbs in the rabbit (Wang et al, 1987), in cattle (MacHugh et al, 1988; Howard et al, 1993a and b), in the pig (Pescovitz et al, 1984; Pescovitz et al, 1990) and in the sheep (MacKay et al, 1985; Dutia and Hopkins, 1991; Bujdoso et al, 1989). Recent investigations of ovine CD1 expression using anti-sheep and anti-cattle mAbs have demonstrated that all the mAbs recognise a 46kD α chain in association with β_2m (S. Rhind, personal communication).

1.6.2 Cellular And Tissue Distribution Of The CD1 Molecules.

The patterns of expression of CD1 molecules have been investigated by immunohistology and FACS analysis in humans, mice, cattle, pigs and sheep. Expression of the various CD1 isoforms appears to be differentially regulated in different tissues. In all species, CD1a, b and c expression is predominantly detected on thymocytes and antigen presenting cells including dendritic cells, Langerhans cells, B cells and monocytes, whereas CD1d expression appears to be mainly localised to intestinal cells.

In humans, CD1 molecules are expressed by virtually all cortical thymocytes, but not by medullary thymocytes nor by peripheral T cells (Amiot et al, 1986; Calabi et al, 1991). The CD1a, b and c molecules are coordinately expressed on the same cells with CD1a levels the highest and CD1c levels the lowest (Calabi and Bradbury, 1991). Expression of CD1a, b and c molecules has also been detected on the thymic lymphoma cell line MOLT4 (Cotner et al, 1981; Kahn-Perles et al, 1985). The significance of CD1 expression in the thymus has been investigated by studying CD1 expression in relation to other cell surface markers such as CD4, CD8 and CD3 (Lanier et al, 1986; Blue et al, 1987; Blue et al, 1989; Gambon et al, 1988). The results indicate that thymocyte maturation is accompanied by a loss of CD1 expression and the simultaneous development of cells into single positive lineages expressing a functional TCR/CD3 complex. Thymic CD1 expression is also inversely correlated to MHC class I expression and it has been suggested that CD1 plays an active role in the developmental processes in the thymus (Calabi et al, 1991).

In addition to cortical thymocytes, human CD1 expression has also been detected on dendritic cells (DC's; Cattoretti et al, 1989) and B cells (Small et al, 1987). Dermal DC's are $CD1a^{+}b^{-}c^{+}$, Langerhans cells (LC's) are $CD1a^{+}b^{-}c^{-}$ and B cells express only CD1c. $CD1^{+}$ DC's can be detected in the T cell rich paracortical regions of lymph nodes, and $CD1^{+}$ B cells have been observed in the marginal zone (MZ) surrounding germinal centres of both lymph nodes and spleen.

Several reports in recent years have demonstrated expression of a fourth CD1 molecule, CD1d, on the majority of intestinal epithelial cells and low level expression on the thymic lymphoma cell line MOLT4 (Bilsland and Milstein, 1991; Balk et al, 1991; Blumberg et al, 1991; Balk et al, 1994). CD1d expressed in the gut may be an important ligand for $CD8^{+}$ T cells (Panja et al, 1993).

The murine CD1d molecules exhibit a similar pattern of expression to that observed in the human situation. Bleicher and colleagues (1990) detected prominent expression of CD1 on murine intestinal epithelial cells and in the cytoplasm of hepatocytes using a rat mAb raised against a murine CD1.1 fibroblast transfectant. This mAb also detected low levels of CD1 expression on thymocytes and peripheral lymph node cells. A rabbit antiserum raised against a recombinant CD1.2 fusion protein detected CD1 expression by Western blot analysis in murine thymus, liver and spleen but not in intestine (Mosser et al, 1991). The discrepancies in detection of intestinal CD1 expression have been resolved by *in situ* hybridisation analysis which demonstrated that only Paneth cells at the bottom of crypts of Lieberkuhn expressed CD1 mRNA at high levels and that no signal was detected in other IEC (Lacasse and Martin, 1992).

The pattern of expression of bovine and porcine CD1 resembles the pattern detected in the ovine system, as described below for the ovine group I mAbs (MacHugh et al, 1988; Howard et al, 1993a; Howard et al, 1993b; Pescovitz et al 1990; Dutia and Hopkins, 1991). Current studies in sheep using a panel of mAbs raised against sheep and cattle thymocytes, dendritic cells and intestinal epithelial cells have provided further data on the extent of ovine CD1

tissue distribution (S. Rhind, personal communication). The mAbs studied cluster into 2 main groups which have differential staining patterns. The mAbs in group I are HCD1b-like and detect CD1 expression on cortical thymocytes, afferent lymph DC's, and DC's in the dermis, lymph node and spleen. The second group of mAbs, exemplified by SBU-T6 (Mackay et al, 1985), have a much wider tissue distribution. In addition to recognising the same cell types as the group one mAbs, CD1 expression is also detected on LC's in the epidermis, on the majority of B cells in the blood, afferent and efferent lymph, on blood monocytes and on cells of dendritic morphology in the lamina propria, liver and lungs.

1.6.3 Molecular Associations Of The CD1 Molecules.

The CD1 α chains associate noncovalently with β_2m. In the case of human CD1a and CD1b molecules, but not CD1c, this association is "loose" since free CD1a and b α chains have been detected at the cell surface (Ziegler and Milstein, 1979; Amiot et al, 1986). Furthermore, extracellular β_2m can exchange with β_2m bound to CD1a α chains but not to CD1c α chains- the CD1a molecules detected on MOLT4 cells grown in tissue culture in the presence of foetal calf serum are preferentially associated with bovine β_2m (Kefford et al, 1984; Bernabeau et al, 1984). It has been suggested that "loose" β_2m association is functionally important and may facilitate intermolecular interactions (Calabi et al, 1991).

It is not known whether association of the classical CD1 α chains with β_2m is an essential requirement for transport to the cell surface, as is the case for MHC class I molecules (Zijlstra et al, 1990). However, recent experiments have indicated that this is not the case for the human CD1d molecule expressed in the intestinal epithelium (Balk et al, 1994). Immunoprecipitations with CD1d-specific mAbs were negative for β_2m, suggesting either cell surface expression in the absence of β_2m, or a loose, easily perturbed, association with β_2m. Transfection of CD1d cDNA into a human β_2m-negative cell line confirmed that CD1d can be expressed at the cell surface in the absence of β_2m.

All three classical human CD1 molecules are additionally found at the cell surface in association with molecules other than β_2m. In particular, covalently associated CD1a/CD8

complexes have been detected on the surface of cortical, but not medullary, thymocytes (Snow et al, 1985; Ledbetter et al, 1985). These disulphide-bonded multimers exist as complexes of 120kD and above, with a minimum stoichiometry of $CD1a_1:CD8_2$. The bond(s) occur in the membrane spanning or cytoplasmic domains of both molecules and β2m is absent from the complexes. The CD1a α chain has also been observed in noncovalent association with CD1b, CD1c and MHC class I α chains at the cell surface of thymocytes but not leukemic cells (Calabi et al, 1991).

1.6.4 Influence Of Cytokines On Expression Of CD1 Molecules.

A series of studies have documented the differential effects of various cytokines on CD1 expression in different cell types (Porcelli et al, 1993; Kasinrerk et al, 1993; Kolenik et al, 1990; Ishii et al, 1990). CD1 expression is inducible by IFN-γ on the human thymoma cell line MOLT4 (Kahn-Perles et al, 1985). Human peripheral blood monocytes, which do not normally express significant levels of CD1a, b or c molecules, were cultured for 60 hours in medium containing 100U/ml of GM-CSF and IL-4 (Porcelli et al, 1993). High levels of expression of the three classical CD1 molecules were detected. Monocytes cultured in the absence of cytokines or with IFN-γ did not express detectable levels of these molecules. A second study of human monocytes found that the expression of CD1a, b, and c molecules was induced by treatment with100ng/ml of rGM-CSF alone (Kasinrerk et al,1993). CD1 expression appeared on day 1 of culture and was upregulated until day 3. Recombinant IFN-γ, TNFα, IL-1α, IL-1β and IL-6 had no effect on CD1 expression. CD1 mRNA was detected by Northern blot analysis in monocytes cultured with rGM-CSF but not in freshly isolated cells. Down regulation of CD1a and CD1c occurred by day 6 of culture but CD1b expression remained stable. The cytokine GM-CSF, known to be produced by keratinocytes, had a totally different effect on human Langerhans cells (Kolenik et al 1990). Culture of LC's with 10ng/ml GM-CSF for 5 days caused a decrease in CD1 expression and an increase in the ability of the cells to stimulate proliferation of allogeneic T cells. The decrease in CD1 expression detected on LC's treated with TNF-α and IL-1 resembled the phenotypic changes observed when LC cells are placed in culture by themselves. In contrast, the cytokines IL-6 and TNF-α, also known to be produced by keratinocytes, induced an upregulation in CD1a expression on gingival LC's

when cultured with cells at 5ng/ml (Ishii et al, 1990). GM-CSF had no effect in this system.

The specific induction of CD1 expression by cytokines on potential antigen presenting cells supports the hypothesis that CD1 molecules are involved in the interactions between accessory cells and T cells.

1.6.5 Relationship Between TL and CD1.

The CD1 molecules were originally considered to be the human equivalent of the mouse TL antigens, on the basis of the pattern of expression and preliminary biochemical evidence (Terhorst et al, 1980; van Agthoven and Terhorst, 1982). However, differences in the peptide backbone sizes of the CD1 molecules compared to the TL molecules gave the first indication that these molecules may not be homologous (van de Rijn et al, 1983). Isolation and sequencing of CD1 genes in the mouse revealed that the CD1 genes are no more closely related to the TL genes than to other class I genes, and RNA transcripts for both TL and CD1 antigens coexist in the mouse thymus (Bradbury et al, 1988). Furthermore, the very limited polymorphism in the mouse does not segregate with MHC haplotype, indicating that the CD1 genes, in contrast to the TL genes, are not encoded in the murine MHC gene complex on chromosome 17. The subsequent mapping of the human CD1 genes to chromosome 1 and the murine genes to chromosome 3 confirmed that the CD1 genes are not MHC encoded and therefore are not homologous to the murine TL genes (Albertson et al, 1988; Bradbury et al, 1991).

1.7. Function(s) Of Class Ib Molecules.

The CD1 molecules are similar in various respects to the nonclassical class I molecules H2Q, T and M in the mouse and HLA-E, F and G in humans. These molecules are collectively referred to as the class Ib proteins, and are characterised by limited polymorphism and restricted tissue distribution in comparison to the classical class I molecules (Stroynowski and Fischer Lindahl, 1994).

Although the structure and expression of the CD1 molecules has been extensively studied in

various species, their function(s) remain ill-defined. That these molecules do have a function is suggested by their structural relatedness to MHC class I molecules and their preferential expression on cells involved in antigen presentation to T cells i.e. Langerhans cells, dendritic cells and B cells. Furthermore, the expression of CD1 in the thymus is tightly regulated during thymocyte differentiation, suggesting a role in thymic developmental processes. CD1 is expressed exclusively on cortical thymocytes at a time when these cells are subject to the selection processes which control MHC restriction and tolerance development. It is striking that CD1 expression is inversely correlated to class I expression on thymocytes (Blue et al, 1989). CD1 has also been shown to be covalently associated with CD8 on thymocytes and there is some evidence that CD1 molecules are linked to a cellular activation pathway (Theodorou et al, 1990). Several investigators have additionally proposed a role for CD1 molecules in antigen presentation to T cells, and in particular to $\gamma\delta$ T cells, as outlined below (Strominger, 1989; Calabi et al, 1991; Janeway et al, 1988). However, it has alternatively been suggested that the class Ib molecules, including CD1, are nonfunctional in comparison to the classical class I molecules, and that their genes represent "rotting hulks in an evolutionary junkyard" (Klein et al, 1991; Lawlor et al, 1990). The reasons cited in support of this view include the fact that nonclassical class I genes in the mouse can be deleted without deleterious effect, the lack of polymorphism which is essential for classical class I function, and the inability to demonstrate an antigen presentation function for these molecules. This view has been challenged in recent years by growing evidence that class Ib molecules can present peptides to both $\alpha\beta$ and $\gamma\delta$ T cells, as outlined in the following sections.

1.7.1 Class Ib Molecules As Ligands For T Cell Receptors.

A role for class Ib molecules in antigen presentation to $\gamma\delta$ T cells was originally proposed on the basis that the limited germline diversity of $\gamma\delta$ TCR's suggested recognition of relatively nonpolymorphic ligands (Janeway et al, 1988; Strominger, 1989). However, this view has been modified since it is now known that the $\gamma\delta$ TCR repertoire is actually larger than the $\alpha\beta$ TCR repertoire due to the great potential for junctional diversity (see section 1.3.2), although the murine $\gamma\delta$ TCR repertoire *is*, in fact, highly restricted at particular epithelial sites (Allison and Havran, 1991). In recent years there has been increasing evidence to suggest that CD4$^-$

CD8⁻ (DN) T cells, irrespective of whether they express $\alpha\beta$ or $\gamma\delta$ TCR's, predominantly recognise nonclassical presentational elements including the class Ib molecules Qa,TL, Blast-1 (TCT.1 or CD48), as well as CD1 molecules themselves. Several studies have reported the involvement of the murine nonclassical class I molecules Qa, TL and HMT in eliciting T cell responses. Three TL-specific murine T cell clones have been characterised in detail. The target ligand for the DN $\gamma\delta^+$ clone G8 has been mapped to the TL region of the MHC using a panel of APC's derived from various inbred mouse strains (Bluestone et al, 1988; Houlden et al, 1989; Matis and Bluestone, 1991). The pattern of reactivity was observed to correlate with a novel polymorphism in the TL locus. The TL-specificity of the DN $\gamma\delta^+$ clone KN6 has been similarly characterised (Bonneville et al, 1989; Matis and Bluestone, 1991; van Kaer et al, 1991) and the gene encoding the specific TLmolecule ($T27^b$, now called $T22^b$) has been cloned (Ito et al, 1990). Milligan and colleagues demonstrated that a short peptide from the influenza virus haemagglutinin protein is recognised by $CD8^+$ $\alpha\beta^+$ cytotoxic T cells in association with a TL-encoded molecule (Milligan et al, 1991). In addition, the TL molecule encoded by the $T3^b$ gene in mice has recently been shown to mediate a transplantation response and elicit $CD8^+$ $\alpha\beta^+$ cytotoxic T cells (Morita et al, 1994). The DN $\gamma\delta^+$ helper T cell clone DGT3, generated by fusion of poly-GT primed lymph node cells with a thymoma cell line, recognises the synthetic copolymer $Glu^{50}Tyr^{50}$ in a Qa-1 dependent manner, and the response is abolished by addition of anti-Qa-1 antibodies (Vidovic et al, 1989). The murine HMT molecules are probably the best defined nonclassical presentational elements (see section 1.3.2) and it has been demonstrated that they can present bacterial peptides to CTL during *Listeria monocytogenes* infection (Pamer et al, 1992; Kurlander et al,1992).

The CD1 molecules themselves have also been implicated in antigen presentation to T cells. The specific recognition of CD1a by the human DN $\alpha\beta^+$ T cell clone BK6 and of CD1c by the human DN $\gamma\delta^+$ T cell clone IDP2 has been demonstrated using CD1 transfected cell lines and blocking antibodies (Porcelli et al, 1989). The J2B7 human DN $\gamma\delta^+$ T cell clone, isolated from the blood of a healthy donor, was found to lyse target cells in a CD1c-dependent manner

(Faure et al, 1992). A human CD8$^+$ $\alpha\beta^+$ intestinal epithelial cell line was found to exhibit CD1 specific cytotoxicity and a dominant clone was CD1c specific (Balk et al, 1991b). Similarly, a report by Panja and colleagues has provided evidence that human CD1d is functionally important in CD8$^+$ T cell activation by intestinal epithelial cells (Panja et al, 1993). More recently, Porcelli and coworkers have reported that the proliferative and cytotoxic responses of a DN $\alpha\beta^+$ T cell line specific for *Mycobacterium tuberculosis* are restricted by the CD1b molecule (Porcelli et al, 1992). Finally, a DN $\gamma\delta^+$ cytotoxic T cell recognising the TCT.1 (Blast-1/CD48) molecule, encoded by a gene in the CD1 region of chromosome 1 in humans, has been described (Mami-Chouaib et al, 1990; Mami-Chouaib et al, 1991; Del Porto et al, 1991).

The studies described above clearly demonstrate that class Ib molecules can be recognised by T cells. However, the majority of the responses actually occur in the absence of any apparent exogenous antigen, and even though most responses are classified as "alloreactive", the clones may be considered to possess autoreactivity due to the nonpolymorphic nature of the class Ib molecules (Porcelli et al, 1991). The requirement for antigen has only been directly demonstrated in four cases- recognition of a synthetic copolymer in the context of Qa-1 (Vidovic et al, 1989), presentation of the influenza peptide by a TL-encoded molecule (Milligan et al, 1991), CD1-dependent T cell responses to mycobacterial peptides (Porcelli et al, 1992), and presentation of bacterial peptides by HMT molecules during *Listeria monocytogenes* infection (Pamer et al, 1992; Kurlander et al, 1992). However, additional evidence has accumulated in recent years to support the idea that class Ib molecules can, in fact, present peptides, as outlined in the following section.

1.7.2 Peptide Presentation By Class Ib Molecules.

Several lines of evidence support the hypothesis that class Ib molecules are able to present peptides (Stroynowski and Fischer Lindahl, 1994). Three types of class Ib molecules have so far been found to depend on the TAP pathway for peptide loading- Qa-1, H2M3 and Qa-2 (Stroynowski and Fischer Lindahl, 1994; Shawar et al, 1994; Aldrich et al, 1992). However, TAP-independent pathways of peptide loading also exist as demonstrated by the fact that

KN6 cells interact with the $T22^b$ molecule on cells deficient in the TAP transporters (Moriwaki et al, 1993). Interaction of class Ib molecules with peptides is further suggested by the ability of a tryptic digest of *Mycobacterium bovis* HSP65 to stabilise cell surface expression of Qa-1 molecules (Imani and Soloski, 1991). Furthermore, mutations in the floor of the putative peptide-binding groove of the $T22^b$ molecule affect recognition by the $\gamma\delta^+$ T cell clone KN6, implicating peptides in this interaction (Moriwaki et al, 1993). Direct interaction of the class Ib molecule Qa- 2 with peptides has recently been demonstrated by analysis of peptides eluted from these molecules. Pool sequencing has indicated that Qa-2 molecules have a more stringent specificity for peptide binding than other class I molecules and that "only a few different peptides are presented" (Rotzschke et al, 1993). However, a more refined analysis of individual eluted peptide sequences revealed that Qa-2 molecules can bind a diverse array of structurally similar peptides in a quantitatively similar manner to the classical class I molecules (Joyce et al, 1994).

1.8 Thesis Objectives.

The ovine CD1 family has previously been investigated only at the protein level using a limited panel of anti-bovine and anti-ovine CD1 mAbs (Hopkins et al, 1989; Bujdoso et al, 1989; Dutia and Hopkins, 1991). The information gained from these studies has not been extensive and all mAbs were found to detect a 46 kD α chain in association with β_2-microglobulin. In order to gain a better understanding of these molecules it was necessary to characterise the ovine CD1 family at the molecular level and correlate the results with biochemical and serological analyses. The primary objective of this study was therefore to characterise the CD1 gene family using well-defined molecular biology techniques, including isolation of CD1 cDNA clones by library screening, and determination of the number of CD1 genes by Southern blot analysis. The ovine CD1 gene probes produced could then be utilised to investigate CD1 expression at the mRNA level in various cells and tissues by Northern blot analysis and *in situ* hybridisation. The availability of gene specific probes would enable the differential expression of different CD1 subtypes in different tissues to be investigated. Characterisation of the CD1 genes and production of a panel of CD1 transfectants would permit correlations to be drawn with the serological data. Overall, definition of the ovine CD1 genes would complement the

existing information from mAb studies and would prove useful in the design of future experiments into CD1 function. The sheep is a particularly suitable experimental animal for this purpose in that CD1 molecules have been implicated in antigen presentation to $\gamma\delta$ T cells, which are a prominent cell type in this species (Mackay and Hein, 1991).

2. Materials and Methods.

2.1 Materials.

The chemicals used during research for this thesis were supplied by Sigma Chemical Co., Poole, Dorset, England, or Fisons Scientific Equipment, Loughborough, England, unless stated otherwise. Radiolabelled compounds were supplied by Amersham International PLC, Amersham, England. Restriction endonuclease enzymes and other DNA modifying enzymes, were purchased from Northumbria Biologicals Ltd. (NBL), Cramlington, U.K., New England Biolabs. (NEB), Hitchin, England, or Boehringer Mannheim, BCL, Lewes, U.K., unless stated otherwise. Recipes for solutions used during this project are given in Appendix1.

2.2 Bacterial Cultures.

2.2.1 Bacterial Growth Media and Plates.

The basic bacterial growth medium was Luria Bertani (LB) broth. This medium was supplemented with the antibiotic ampicillin for selection of transformed cells, since all plasmid vectors used during this project contained the gene which conferred ampicillin resistance. For blue/white selection (Sambrook et al,1989) of bacteria transformed with pTZ or pBluescript plasmids, LB/agar plates were made with LB/1.5% (w/v) bacteriological agar (Oxoid, Basingstoke,England), supplemented with 100μg/ml ampicillin, and 0.75%(w/v) top agar supplemented with 100μg/ml ampicillin, 170μg/ml isopropylthio-β-D-galactoside (IPTG-Melford Laboratories Ltd., Ipswich, England), and 200μg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactoside (Xgal-Melford Laboratories Ltd.).The plasmid pCR^{TM}II from the TA Cloning kit (Invitrogen Corporation, San Diego, California) was transformed into INVαF' competent *E.coli*. In this system, transformed cells were selected on plates containing ampicillin and X-gal only.

2.2.2 Bacterial Strains, Stocks and Culture.

E.coli strain JM101 (F' traD36 lacI^q Δ[lacz]M15 proAB/supE thi Δ[lac-proAB]) (Yannish-Perron et al, 1985)was used for bacterial transformations involving pTZ or pBluescript

plasmids. The *E.coli* strain Y1090 (F⁻ Δ[lac]U169 lon-100 araD139 rpsL[Str^r] supF mcrA trpC22::Tr10 [pMC9;Tet^{r}Amp^r] (Huynh et al, 1985) was used as the host for plating out λgt11 cDNA libraries. Single colonies of untransformed *E.coli* were grown by streaking out cells from frozen stocks onto agar plates made with M9 minimal medium for JM101 cells, and LB/Amp for Y1090 cells. Transformed JM101 cells were grown on LB/Amp plates. Bacterial stocks were frozen in 20%(v/v) glycerol and stored at -70°C.

2.2.3 Preparation of Competent Bacteria.

Competent *E.coli* cells, ready for transformation with recombinant plasmid, were prepared by two slightly different modifications of the method of Hanahan (1983), as follows.

Method 1

0.2ml of a saturated overnight culture of *E.coli* grown in LB medium, was diluted into 20ml of Psi broth and grown with shaking at 37°C until the optical density at 550 nm (OD_{550}) was 0.3. 5ml of this culture was made up to 100ml in Psi broth and grown until the OD_{550} was 0.48. Cells were briefly cooled on ice and pelleted by centrifugation at 2,000xg, 4°C for 5 minutes. The cell pellet was resuspended in 33ml ice cold transformation buffer I (TfBI) and placed on ice for 15 minutes. The cells were pelleted as before, resuspended in 4ml transformation buffer II (TfBII), and placed on ice for 20 minutes. Aliquots (0.2ml) were then transferred to eppendorf tubes, snap frozen in liquid nitrogen and stored at -70°C until required.

Method 2

1ml of a saturated overnight culture of *E.coli* grown in LB medium was diluted into 100ml of fresh LB medium, and grown at 37°C with shaking until the OD_{550} was ~ 0.5. Cells were pelleted from 10ml of this culture by centrifugation at 1,300xg, 4°C, for 5 minutes, resuspended in 800ml transforming buffer (TFB) solution and placed on ice for 5 minutes. Cells were maintained on ice while all further solutions were added- dimethylsulphoxide (DMSO) to 3.5% for 5 minutes, dithiothreitol (DTT) to 75mM for 5 minutes, and a final addition

of DMSO to 7.0%. The competent cells were kept on ice and used as soon as possible, 100µl per transformation reaction.

2.2.4 Transformation Procedure.

Competent bacteria were thawed, or maintained, on ice, and incubated with the ligated DNA (10µl) on ice for 30-60 minutes. Cells were heat shocked for 90 seconds at 42°C. LB medium (800µl) was then added and the cells incubated at 37°C for ~ 60 minutes, to allow expression of the ampicillin resistance gene. Cells were then gently pelleted (10 second spin in microfuge, 11,000xg), resuspended in 100µl of LB medium, and plated out onto plates containing appropriate antibiotic and chemicals for blue/white colour selection. Plates were incubated at 37°C for 16 hours, and single white colonies picked for plasmid preparations.

2.3 Preparation of Plasmid Vectors and Inserts for Cloning.

2.3.1 Preparation of Plasmid Vector for Cloning.

2-5µg of pBluescript plasmid vector (Stratagene, La Jolla, California) was digested with 5-20 Units of enzyme(s) in a total volume of 20µl, using the manufacturer's buffer. Single enzyme digest controls were included where double enzyme digests were carried out. Following digestion, the 5' end of the linearised plasmid was dephosphorylated to help prevent religation by addition of 1-2 Units of calf intestinal phosphatase (CIP) (Boehringer Mannheim), and the manufacturer's buffer. Reactions were incubated at 37°C for 30 minutes followed by addition of 1 Unit of CIP for further 5 minutes at 37°C. Inactivation of CIP was achieved by addition of 1/20 volume of 0.5M EDTA pH8.0, and heating to 65°C for 10 minutes. CIP inactivation was completed by extraction with phenol/chloroform and chloroform. Following ethanol precipitation, the DNA pellet was resuspended in 10µl SDW or TE and analysed by gel electrophoresis.

2.3.2 Preparation of DNA Fragments for Blunt Ended Cloning.

DNA fragments for blunt ended cloning were prepared using the Klenow fragment of *E.coli* DNA polymerase I (Stratagene, La Jolla, California) to " fill in " recessed 3' ends, and T4

polynucleotide kinase (NEB) to phosphorylate the 5' ends. The DNA fragments were incubated at 37°C for 30 minutes with ATP and 10 Units of kinase in the manufacturer's buffer in a total volume of 50µl. Approximately 10-20 Units of Klenow were then added along with 100µM of each dNTP and a further 1µl of 10xbuffer (total volume 60µl), and the incubation continued for a further 30 minutes. Phenol/chloroform and chloroform extractions were performed, followed by ethanol precipitation, and the DNA fragments resuspended in 10-20µl SDW in preparation for cloning.

2.4 Ligation of DNA Fragments.

Ligations were generally carried out using 70-100ng of vector DNA with the insert DNA present at threefold molar excess. Ligations were performed in 10µl volumes containing 5 Units of T4 DNA ligase (NBL, Cramlington, U.K.) and the manufacturer's buffer, and incubated at 4°C for 16 hours. Appropriate controls were included to assess ligase, CIP and transformation efficiency.

2.5 Restriction Endonuclease Digestion of DNA Samples.

Restriction endonuclease digests of DNA, for example λ bacteriophage DNA or pBluescript plasmid DNA, were carried out in a total volume of 20µl using the manufacturer's buffer. Typically, 5µl of DNA from λ minipreps., or 3µl DNA from plasmid minipreps., were digested with 5-20 Units of enzyme(s) for 1-2 hours at 37°C. Digested samples, along with uncut or singly cut controls, were analysed by gel electrophoresis.

2.6 Purification and Quantitation of DNA.

DNA samples were routinely extracted with equal volumes of phenol/chloroform and chloroform, and precipitated in a dry ice/ethanol bath with 1/10th volume 0.3M sodium acetate and 2 volumes of 100% ethanol. Precipitated DNA was pelleted at 11,000xg for 10 minutes at 4°C, washed with 70% ethanol, briefly air-dried, and resuspended in TE or SDW prior to further manipulation. Alternatively, ammonium acetate, pH7.5 (final concentration 2-2.5M), was used in the precipitation step when it was necessary to purify the DNA from high

concentrations of excess nucleotides in the sample.

Following plasmid DNA isolation and restriction endonuclease digestion to obtain insert fragments or linearised vector, DNA was analysed on a 0.8-1.5% (w/v) agarose gel containing 1μg/ml ethidium bromide. TAE buffer was used when it was necessary to subsequently purify DNA from the gel, since TBE buffers inhibit efficient DNA recovery from gels using the Geneclean II system (Stratech Sci.).TBE buffer was used on all other occassions. Before loading, samples were mixed with 1/6 volume of x6 DNA sample buffer. Gels were run at 85-100V for 60-90 minutes, at constant voltage. EcoRI/HindIII digested lambda DNA molecular weight markers (NBL, Cramlington, U.K.) were included on every gel.

Two methods for isolating DNA fragments from agarose gels were employed. Fragments greater than 0.5kb in size were extracted using the Geneclean II kit (Stratech Sci.) according to manufacturer's instructions. Fragments less than 0.5kb in size are not recovered at optimal efficiency using this kit. Spinbind DNA extraction units (FMC Bioproducts, Rockland, Maine, USA) were used to extract DNA fragments less than 0.5kb, according to the manufacturers' protocol.

The concentration of DNA in a purified sample was generally estimated by running a 10% aliquot on a 0.8-1.5% (w/v) agarose gel, and examining the intensity of fluorescence under UV illumination. The concentration of DNA in larger preparations was measured spectrophotometrically by determining the absorbance at a wavelength of 260 nm (OD_{260}). An OD_{260} of 1.0 is equivalent to 50μg/ml double stranded DNA, or 40μg/ml single stranded DNA or RNA. The OD 260/280 ratio gave an estimate of the purity of the sample, where a ratio of 1.8 indicated a sample free from protein or phenol contamination (Sambrook et al, 1989).

2.7 Preparation of ^{32}P-labelled Probe.

DNA fragments to be used as probes were gel purified and the DNA extracted from gel slices using Spinbind DNA extraction units (FMC Bioproducts) as previously described (section

2.6). The probes were labelled with α^{32}P-dCTP (Amersham, specific activity 400Ci/mmol.) using the Pharmacia oligolabelling kit (Pharmacia LKB Biotech.) according to the manufacturer's protocol. This method, based on that of Feinberg and Vogelstein (1983,1984), routinely labelled 50ng quantities of DNA to 10^8-10^9cpm/μg DNA, as determined by acid precipitation on glass fibre discs with 5% trichloroacetic acid (TCA) (Sambrook et al, 1989). Dried discs were placed in 2ml Optiphase Scintillant (Pharmacia LKB Biotech.), activity measured in an LKB Rackbeta Liquid Scintillation Counter, and the specific activity calculated. Only probes with a specific activity greater than 10^8cpm/μg DNA were used in hybridisations. Spinbind DNA extraction units were used to purify the labelled probe from excess nucleotides and free label. Immediately prior to use, probe DNA was denatured by boiling for 10 minutes and quenching on ice.

2.8 Preparation of M13 Replicative Form DNA.

The human CD1 probes were provided as M13 DNA at 20ng/μl in TE buffer. Double stranded DNA for use as a probe in library screening was obtained by preparation of M13 'phage stock and large scale preparation of M13 replicative form DNA, as described below.

2.8.1 Preparation of M13 'Phage Stock.

A titration of the M13 DNA was carried out as follows. Competent JM101 *E. coli* were prepared (2.2.3), and a series of dilutions of the M13 DNA made up in TE, such that 1μl volumes contained 2ng, 0.2ng and 0.02ng of the M13 DNA. The competent bacteria were transformed with 1μl, 2μl and 5μl volumes of each M13 dilution (2.2.4), mixed with 0.9ml LB/20mM glucose and incubated for 30 minutes at 37°C. Transfected cells were mixed with 3ml top agar, IPTG added to 170μg/ml, X-gal added to 200μg/ml and the cells plated out onto LB plates. The plates were incubated overnight at 37°C. This procedure is estimated to produce ~1000 plaques per ng of DNA transfected (Sambrook et al, 1989).

One plaque was picked from a plate containing well separated plaques, deposited into 1ml of SM buffer, and left at room temperature for 1-2 hours to allow the 'phage to diffuse out of the

agar plug. 100µl of the 'phage suspension was mixed with 2ml of plating bacteria (one JM101 colony grown in 50ml LB for 6-8 hours, but not to saturation) and incubated at 37°C for 5-6 hours with agitation. 1ml of this culture was centrifuged in an eppendorf at 11,000xg for 5 minutes at room temperature. The supernatant containing the ss form of the 'phage was retained as the M13 'phage stock.

2.8.2 Large Scale Preparation of M13 Replicative Form DNA.

2.5ml of plating bacteria (JM101) were inoculated with 0.5ml of 'phage stock, mixed by tapping the tube, and incubated at room temperature for 5 minutes. This culture was diluted into 250ml prewarmed LB in 2 litre flasks and incubated at 37°C for 15 minutes with agitation. Chloramphenicol was added to a final concentration of15µg/ml and the culture incubated for a further 2 hours at 37°C. Chloramphenicol inhibits bacterial replication and allows the replicative form of the 'phage to accumulate inside the cells (Sambrook et al,1989). Cells were harvested by centrifugation at 4000xg for 15 minutes at 4°C, the supernatant removed and the pellet resuspended in 100ml ice-cold STE. The cells were harvested as before and the M13 circular replicative form DNA isolated by alkaline lysis (see 2.10.1 and Sambrook et al,1989).

2.9 Screening of cDNA Bacteriophage Libraries By Colony Hybridisation

2.9.1 Library Plating and Nitrocellulose Lifts.

Approximately $2.4x10^5$ plaques were screened from each cDNA library, at~$3x10^4$ pfu per plate. The titre of each library had previously been determined by serial dilution of the library 'phage stock and plating out in exactly the same manner as outlined here. An appropriate number of 'phage were added to a mixture of 3ml adsorption buffer (10mM $CaCl_2$/10mM $MgCl_2$) and 2ml of host Y1090 *E.coli* grown to saturation, usually overnight, in LB broth supplemented with 0.4% (v/v) maltose. This mixture was incubated for 10-15 minutes at 37°C to allow the bacteriophage to adsorb to host cells. The adsorbed 'phage mix was added to an appropriate volume of LB/agar (0.75%, w/v) supplemented with 10mM magnesium sulphate and 0.6% glucose, and plated out onto LB/agar (1.5% w/v) plates containing10mM

magnesium sulphate. Plates were incubated at 37°C for 16 hours.

After incubation, the 'phage plates were cooled at 4°C for 30 minutes. Nitrocellulose filters (Immobilon-NC membrane, Millipore,UK) were numbered and labelled using pencil, and carefully placed, numbers down, onto the cooled plates. Filters were left in place for 5 minutes, and alignment holes made using a sterile 18-gauge needle. Meanwhile, four pads of Whatman 3MM paper were prepared, each three layers thick. Two pads were soaked with alkaline solution (0.2M NaOH) and two with neutralizing solution (0.5M Tris pH 7.5, 1.5M NaCl), and excess fluid was poured off. Filters were removed from the plates and sequentially placed, numbers up, on each alkaline and neutralizing pad for 2 minutes each. Filters were dried at 37°C, then baked for 2 hours under vacuum at 80°C to fix the DNA to the filters.

2.9.2 Hybridisation of DNA Probes to Nitrocellulose Lifts

Baked nitrocellulose filters were prehybridised for 4-6 hours in a minimum volume of hybridisation solution, but ensuring that the filters could move independently of each other. After prehybridisation, the denatured radioactively labelled probe was added to the hybridisation solution. Heterologous hybridisation reactions were performed at 58°C for 16 hours and homologous hybridisations at 65°C for 16 hours.

Following overnight hybridisation, the filters were washed as follows. For both heterologous and homologous hybridisations, an initial rinse with 2xSSC/0.1%SDS was carried out for 2-3 minutes at room temperature. Filters used for heterologous hybridisations were then washed for 30-60 minutes in 0.5% SSC/0.1%SDS at 58°C. Filters used for homologous hybidisations were successively washed for 30-60 minutes in 0.5xSSCX/0.1%SDS at 65°C and 20-60 minutes in 0.2xSSC/0.1%SDS at 65°C. In each case the radioactivity coming off in the wash solution and that remaining on the filters was monitored at regular intervals, and washing stopped when very little radioactivity appeared to be coming off the filters. Filters were air-dried and then exposed to Kodak X-OMAT R or S film, with a phosphotungstate

intensifying screen, at -70°C for 1-3 days.

2.9.3 Secondary Screening of Bacteriophage Libraries.

Positive plaques from primary screens were picked into 1ml SM buffer and 3 drops of chloroform to form a 'phage stock. A 1/500 dilution of these 'phage stocks was used to infect host cells for secondary screening, and plated out essentially as described for primary screening. Plates which grew dense plaques or too few plaques at this dilution were not screened. Rather, alterations were made to the 'phage dilution to give an appropriate density of plaques for secondary screening. Nitrocellulose lifts were taken and hybridisations and washes carried out as previously described. Positive plaques were picked for preparation of λ DNA, as follows.

2.9.4 Preparation of Bacteriophage Lambda (λ) DNA

Following identification of CD1 positive plaques in secondary screens of various λgt11 cDNA libraries, small scale preparations of λ DNA were carried out. Positive plaques were picked into a mixture of 0.3ml adsorption buffer (10mM $CaCl_2$/10mM $MgCl_2$) and 0.2ml of dense overnight culture of host Y1090 *E.coli* grown in LB/0.5% maltose. This mixture was incubated at 37°C for 10-15 minutes to allow phage adsorption to host cells. 10ml of LB/10mM $MgSO_4$/0.6%(v/v) glucose was added, and the cultures incubated with shaking at 37°C for 16 hours. The cells were then pelleted by centrifugation at 1,300xg for 5 minutes, The supernatant containing the λ 'phage was then centrifuged at 150,000xg for 30 minutes at 4°C, and the 'phage pellets resuspended in 100μl SM buffer. Proteinase K (Sigma), made up in SM buffer, was added to a final concentration of 0.5mg/ml and the mixtures incubated at 37°C for 2 hours. The samples were extracted with equal volumes of phenol/chloroform and chloroform, followed by ethanol precipitation. The DNA was pelleted (10 minutes, 11,000xg), washed in 70% ethanol, briefly dried,and resuspended in 50μl TE. 5μl of each sample was digested with the appropriate enzyme and examined by gel electrophoresis for the presence of excisable inserts.

2.10 DNA Sequencing.

Both double stranded and single stranded DNA sequencing were carried out based on the dideoxy chain termination method of Sanger et al (1977).

2.10.1 Small Scale Preparation of Plasmid DNA for Double Stranded Sequencing.

Transformed *E.coli* were grown at 37°C for 16 hours with shaking in 12-25ml LB broth containing 10mM $MgSO_4$ and 0.6% (w/v) glucose. Plasmid DNA was isolated by a modification of the methods of Ish-Horowicz and Burke (1981) and Birnboim and Doly (1979). Bacterial cells were pelleted (1,300xg, 4°C, 10 minutes), resuspended in 200µls of GTE (solution I), and lysed by the addition of 400µl of freshly made alkaline/SDS solution (solution II). Addition of 300µl potassium acetate solution (solution III) precipitated chromosomal DNA. This was removed, along with protein/SDS complexes and high molecular weight RNA by pelleting in a microfuge (11,000xg, 5 minutes, room temperature). The remaining RNA was removed by incubation in 50µg/ml DNAase free RNAase A (Pharmacia) at 37°C for 1-2 hours. Phenol/chloroform and chloroform extractions were performed, and the DNA precipitated by addition of 500µl (0.6 volumes) isopropanol. The precipitated DNA was pelleted (11,000xg, room temperature, 10 minutes), washed with 70 % ethanol, dried briefly,and resuspended in 50µl TE. 3µl of this sample was used for restriction endonuclease digestion and gel electrophoresis, and 5µl (~3-5µg DNA) for sequencing. The plasmid DNA isolated by this method is sufficiently pure for double stranded DNA sequencing to be carried out. Typical DNA yields from these preparations were in the range 3-5µg plasmid DNA per ml of original culture (Sambrook et al, 1989).

2.10.2 Large Scale Preparation of Plasmid DNA for Double Stranded Sequencing.

The method used for preparation of plasmid DNA from 100ml cultures was essentially the same as described for 12ml preparations, but using proportionately larger solution volumes. For 500ml cultures, plasmid DNA was prepared using pZ523 columns (5 Prime-3 Prime Inc., West Chester, USA), according to the manufacturers protocol. The plasmid DNA isolated by this method was sufficiently pure to allow double stranded sequencing to be carried out. The initial part of this protocol (bacterial cell lysis) was essentially a scaled up version of the alkaline

EDINBURGH UNIV. COL. EX LIBRIS

lysis method described above (section 2.10.1). RNA was then removed by incubation with RNAase at 50μg/ml for 1-2 hours at 37°C. Phenol/chloroform and chloroform extractions were performed. The aqueous layer was removed, 360μl of 5M NaCl added, and the volume adjusted to precisely 1.8ml, using TE. The pZ523 column was spun at 1100xg for 1 minute in a swinging bucket rotor to remove the storage buffer. The 1.8ml sample was carefully loaded onto the column, and the column spun at 1100xg for 12 minutes in the swinging bucket rotor. The volume of collected effluent was measured, and 0.6 volumes of 100% isopropanol added to precipitate the DNA. The DNA was pelleted at 16,000xg for 20 minutes at room temperature and the pellet washed three times with 70% ethanol. The pellet was resuspended in 250μl of TE or SDW prior to sequencing, and analysed by restriction endonuclease digestion and gel electrophoresis to estimate the quantity and purity of the DNA sample.

2.10.3 Preparation of Plasmid DNA for Single Stranded Sequencing.

Single stranded plasmid DNA for sequencing was basically prepared according to the Pharmacia protocol. Transformed *E.coli* were grown at 37°C with shaking in 2xYT broth supplemented with ampicillin (150μg/ml), until the OD_{660} was ~0.5-0.7. 400μl of this culture was infected with the helper 'phage M13K07, at a multiplicity of infection (MOI) of 10, and cultured at 37°C with shaking for 1 hour. The culture was then made up to 10ml with 2xYT supplemented with ampicillin (150μg/ml) and kanamycin (70μg/ml), and grown at 37°C with shaking for 16 hours. Bacterial cells were pelleted by centrifugation (2,000xg, 4°C, 10 minutes) and 'phage were precipitated from the supernatant by the addition of 2.5ml polyethylene glycol solution (20%PEG, 3.5M NaCl) on ice for 30 minutes. Precipitated 'phage were pelleted at 11,000xg for 40 minutes at 4°C, and the pellet resuspended in 500μl STE solution. This solution was extracted twice with phenol/chloroform, once with chloroform, and the DNA ethanol precipitated. The single stranded DNA was resuspended in 30μl SDW, and ~7μl used for each sequencing reaction.

2.10.4 Sequencing Reactions.

Both double and single stranded sequencing were carried out using the Sequenase Version 2.0 kit (USB Corporation) with the supplied Sequenase Version 2.0 T7 DNA polymerase, and α-^{35}S-dATP (Amersham: specific activity 400Ci/mmol.) according to the manufacturer's protocol. Prior to double stranded sequencing, the DNA template was denatured by treatment with alkali. Approximately 3-5μg of plasmid (5μl from miniprep. DNA) was placed in an eppendorf tube and the volume raised to 20μl with SDW. 2μl of denaturing solution (2M NaOH/ 5mM EDTA) was added, mixed well, and the solution left at room temperature for 5 minutes. Denatured DNA was precipitated in a dry ice/ethanol bath by addition of 6.2μl 3M sodium acetate, 7μl SDW and 75μl 100% ethanol. The precipitated DNA was pelleted by centrifugation (11,000xg, 10 minutes, 4°C), washed with 70% ethanol, and resuspended in 7μl SDW. The majority of sequencing reactions were carried out in the presence of 10% dimethylsulphoxide (labelling mix diluted 1/5 in 50% DMSO), to help overcome problems with compressions.

2.10.5 Sequencing Gel Electrophoresis.

DNA sequencing gels were composed of 6% or 8% (w/v) acrylamide (29:1 ratio of acrylamide:N-N methylene bisacrylamide [Biorad]), 8M urea (Sigma; molecular biology grade), and 1xTBE-A buffer. Deionised formamide (10% v/v) was sometimes added to the gel mix, as this chemical helps to destabilise any secondary structures in the DNA template which may cause compressions (USB,1990). The sequencing plates were sealed at the base using the casting tray with 60ml acrylamide, 200μl TEMED and 400μl of 25%(w/v) AP (Biorad). The acrylamide solution (120ml) was degassed and filtered through a 0.45μm filter prior to adding 100μl TEMED, 400μl of 25% AP, and casting the gel. The gel casting plates and sequencing tank equipment were either from Hybaid Ltd.,U.K. or Biorad Sequi-Gen. The running buffer used in the upper and lower reservoirs was 1xTBE-M.

Just prior to sample loading, urea was washed out of the wells using a needle and syringe. The sequencing samples were heated to 75°C for 2 minutes, and 1.5-3.0μl of each sample was quickly loaded onto the gel. For short runs (3-4 hours), the gels were prerun for 30

minutes to 1 hour until the temperature reached 45-50°C. Gels were run for 3-10 hours at 60 Watts. Sequencing gels were dried down on Whatman no.3 paper at 80°C under vacuum on a Biorad model 583 gel drier. Gels were exposed to Kodak X-OMAT R or S film for 1-3 days at room temperature.

2.10.6 Analysis of Sequence Data.

Sequence data was analysed using version 7.3 of the University of Wisconsin Genetics Computer Group Sequence Analysis software (UWGCG 7.3; Devereux et al,1984).

2.11 Southern Blot Analysis

2.11.1 Preparation of Genomic DNA.

Genomic DNA was prepared according to the methods of Enrietto (1983) and Gross Bellard et al (1972). Efferent lymph cells were used as a readily available source of tissue for genomic DNA preparation. Cells from 12 samples, each containing 1.5×10^7 cells, were pelleted by centrifugation at 500xg for 5 minutes, and washed twice in ice-cold PBS. Pelleted cells were resuspended in 400µl per sample of lysis buffer and incubated at room temperature for 15 minutes. Proteinase K was added to a final concentration of 100mg/ml, and incubated at 68°C for 1 hour. Lysed, digested cells were extracted once with an equal volume of phenol/chloroform pH8.0 at 68°C, once with phenol/chloroform at room temperature, and once with chloroform only at room temperature. Nucleic acids were precipitated in a dry ice/ethanol bath by addition of sodium acetate (0.15M final concentration) and 2 volumes of 100% ethanol. Precipitated nucleic acids were pelleted at 1700xg for 10 minutes, washed in 70% ethanol, dried briefly and resuspended in 400µl TE containing 50mg/ml RNAase. Digestion of RNA was carried out at 37°C for 1 hour. Genomic DNA was extracted once with phenol pH8.0, twice with phenol/chloroform and twice with chloroform/IAA. DNA was precipitated with sodium acetate and 100% ethanol as before. Pellets were washed in 70% ethanol, briefly airdried, and resuspended in 200µl TE. To facilitate solubilization, the DNA was stirred gently at 4°C for 16 hours. The final concentration of DNA was determined spectrophotometrically.

2.11.2 Southern hybridisation.

Approximately 10μg samples of genomic DNA were digested with 15-45 Units of a variety of restriction endonucleases in the manufacturer's buffer, for 16 hours at 37°C. A further 5-15 Units of enzyme was added to each sample, mixed thoroughly and incubated for 6 more hours at 37°C. Digested DNA was run on a 0.8% (w/v) agarose gel in TAE buffer for 16-18 hours at 40V, constant voltage. The gel was examined under UV light to check for completion of digestion, and a photograph taken.

Following electrophoresis, DNA was transferred to a nylon membrane (Hybond N, Amersham) according to a modification of the protocol of Southern (1975). After electrophoresis, the DNA was first depurinated by soaking the gel with gentle shaking in 0.25M HCl for 15 minutes, and denatured/neutralised by soaking in 1.5M NaCl /0.5M NaOH for 30 minutes. Blotting of DNA onto the nylon membrane was carried out overnight as described by Sambrook et al (1989) using alkaline transfer buffer (0.4 M NaOH). After blotting, the gel was examined under UV light to check the efficiency of DNA transfer to the filter. The nylon filter was rinsed in 2xSSC for 5 minutes, and allowed to airdry.

2.11.4 Hybridisation of Nylon Filters.

Nylon filters were prehybridised for 2-4 hours at 42°C in a formamide containing prehybridisation solution. The ^{32}-P labelled DNA probe (see section 2.7 for labelling procedure) was denatured by boiling for 5-10 minutes, added to a minimum volume of prehybridisation solution, and the incubation continued for a further 16-18 hours at 42°C. Nylon filters were then rinsed in 2xSSC/0.1%SDS for 5 minutes at room temperature with gentle agitation, followed by 20-30 minutes at 42°C in 0.2xSSC/0.1%SDS and 20-30 minutes at 42°C in 0.1xSSC/0.1%SDS. Filters were air-dried at room temperature, and exposed to Kodak X-OMAT S film in a casette with an intensifying screen, at -70°C for 5-15 days.

2.12 RNA Isolation.

Both total RNA and polyA+ RNA were isolated for use as templates in cDNA synthesis reactions. The precautions taken to minimise RNase contamination included treatment of SDW for making up solutions with 0.2% diethylpyrocarbonate (DEPC) followed by autoclaving, baking glassware at 120°C overnight, and wearing gloves at all times (Sambrook et al,1989).

2.12.1 Isolation of PolyA+ RNA.

The FastTrack mRNA Isolation kit (Invitrogen Corporation, San Diego, California) was used to isolate polyA+ RNA. Tissue removed at postmortem was immediately snap frozen in liquid nitrogen, and stored at -70°C for a short time before use. Approximately 1g of frozen tissue was cut into small pieces and ground in a pestle and mortar with a little dry ice. The crushed tissue was placed in a sterile, baked Dounce homogenizer and the polyA+ RNA isolation continued according to the manufacturer's protocol. Briefly, the tissue was lysed in a detergent-based buffer containing RNase and protein degraders for 60 minutes at 45°C. High molecular weight DNA was thoroughly sheared by passing the lysate through a sterile plastic syringe fitted with a 21-gauge needle until it became relatively easy to do this. The salt concentration was adjusted and the lysate applied to oligo(dT) cellulose for adsorption. The DNA, proteins and cell debris were washed off in low salt buffer, the mRNA eluted in elution buffer and polyA+ RNA precipitated from the eluate by ethanol precipitation with sodium acetate. The RNA pellet was resuspended in 20-50µl elution buffer, and the yield determined by spectrophotometry (an OD_{260} of 1.0 is equivalent to 40µg/ml of RNA).

2.12.2 Isolation of Total RNA.

Total RNA was isolated by single step extraction according to the method of Chomczynski and Sacchi (1987).

Small pieces of tissue removed at postmortem were immediately snap frozen in liquid

nitrogen and stored briefly at -70°C until use. Approximately 200mg of frozen tissue was cut into small pieces and placed in a Dounce homogenizer. 2ml of GTC (guanidinium isothiocyanate) buffer was added and the tissue quickly homogenized. The lysate was mixed well with 200µl sodium acetate (pH 4.0), 400µl chloroform and 2ml water-equilibrated phenol, aliquoted into eppendorf tubes and placed on ice for at least 15 minutes. The mixture was then centrifuged at 11,000xg for 10 minutes at 4°C. 400µl of aqueous supernatant was removed, mixed with an equal volume of isopropanol and placed on ice for 10 minutes. Precipitated RNA was pelleted by centrifugation (11,000xg, 10 minutes, 4°C).The supernatant was discarded and the pellet washed twice in 75% ethanol. The RNA pellet was air-dried and resuspended in 100µl of DEPC treated sterile water.

2.12.3 Gel electrophoresis of Total RNA.

Small aliquots of total RNA were separated by electrophoresis through a formaldehyde-containing gel to check the integrity of the samples prior to further manipulation. The 1.5%(w/v) denaturing agarose gel was prepared as follows. 4.5g of agarose was mixed with 214ml sterile DEPC- treated water and 30ml 10xMOPS (3-[N-morpholino]-propanesulfonic acid) buffer. This was boiled to dissolve the agarose, cooled to 60°C, and 54ml formaldehyde added in a fume hood just prior to pouring the gel. RNA samples were prepared by mixing 5µl of sample with 15µl of RNA loading buffer, and heating to 70°C for 15 minutes. One fifth volume (4µl) of RNA dye mix was added, and the samples loaded onto the gel. The gel was run at 65V constant voltage in a cold room for 1.5-2 hours. The gel was examined under UV light to check the samples. Samples containing degraded RNA were visible as smears lacking clear 28S and 18S ribosomal RNA bands, and were discarded.

2.13 cDNA Synthesis and Tailing Reactions.

2.13.1 cDNA Synthesis

First-strand cDNA synthesis from either total RNA or poly A^+ RNA was carried out using the Amersham cDNA Synthesis Kit (cat.no. RPN1256Z) and random hexanucleotide primers according to the manufacturer's protocol. The reverse transcriptase enzyme in this kit was

derived from the Rous associated virus 2 (RAV2). The efficiency of the synthesis reaction was estimated by inclusion of α-^{32}P dCTP in the reaction mix and measurement of the specific activity of the TCA precipitable fraction, as described previously (see section 2.7). Oligonucleotide primers and excess dNTP's were removed by ammonium acetate precipitation of the cDNA (carried out twice), and the cDNA pellet resuspended in either 30.6µl of TE for subsequent tailing reactions, or 400µl of TE for use as a PCR template (10-15µl used per PCR reaction).

A similar protocol was used to synthesis cDNA for the PCR investigations of gut CD1 expression. Briefly, cDNA was synthesised from total RNA in a 20µl reaction mix composed of Superscript buffer (Gibco BRL), dNTP's, random hexanucleotide primers, DTT and Superscript enzyme (Gibco BRL). This enzyme is derived from a cloned Moloney Murine Leukemia virus (M-MLV) gene. The reaction was carried out at 45°C for 45 minutes and terminated by heating to 95°C for 5 minutes. The cDNA mixture was cooled on ice, diluted to 40µl inTE, and 3µl used as template in each PCR reaction.

2.13.2 Tailing Reactions.

The enzyme terminal deoxynucleotidyl transferase (Boehringer Mannheim) was used to tail the cDNA's with either poly G or poly A, according to the manufacturer's protocol and the method outlined by Thiessen et al (1990). This protocol resulted in addition of 15-25 residue tails on the cDNA at the 3' end. The enzyme was inactivated by heating to 65°C for 5-10 minutes. The cDNA pool was diluted into a final volume of 200-400µl with TE, and 10-15µl used as a DNA template in each PCR reaction.

2.14 Polymerase Chain Reaction (PCR).

PCR reactions were carried out in 50 or 100µl volumes in sterile 0.5ml eppendorf tubes. The basic PCR reaction components were assembled as follows- 1xPCR reaction buffer (50mM KCl, 10mM Tris.Cl pH8.3, 1.5mM $MgCl_2$), 100µM each dNTP (Pharmacia), 0.17mg/ml bovine serum albumin (BSA), DNA template, 25-50pmoles of each primer and 2 units of *Thermus*

aquaticus (Taq) polymerase enzyme. Ohara's buffer contained 3mM DTT in addition to the basic reaction buffer components (Ohara et al,1989). The various PCR reaction mixes and cycling parameters used during this project are detailed in Table 1. When all the reaction components were assembled, a mineral oil overlay was added and cycling carried out as described, using either a Techne Programmmable Driblock PHC-1 or a Hybaid Omnigene thermal cycler. Following completion of PCR amplification, 10% of the product was analysed by gel electrophoresis, as described previously (2.6).

In some instances, additional components were added to the reaction mix, as indicated in the table. DMSO is a denaturant and is thought to contribute to PCR amplification success by reducing the secondary structure of the DNA template, when used at concentrations of 10% (v/v) or less (Pomp and Medrano,1991: Rolfs et al, 1992). The addition of the detergent w1 is thought to contribute to the maintenance of enzyme stability during amplification, and its inclusion depended on the source of the enzyme used. For some PCR reactions, the reaction tubes were heated to 94°C for 4 minutes and cooled to 80°C prior to addition of the Taq. polymerase enzyme ("hot-start"), as this has been shown to improve the yield and specificity of the PCR product (D'Aquila et al, 1991).

The primers used for PCR were synthesised by Oswel DNA services (Dept. of Chemistry, Kings Buildings, West Mains Road, Edinburgh, EH9 3JJ) and their sequences are given in Appendix 2. Typically, PCR primers were18-34 nucleotides in length and ~50% GC rich so that both primers in a reaction had similar optimal annealing temperatures. Ideal primers contained a random distribution of bases. Primer sequences were selected to avoid those which could potentially form hairpin loops. It was particularly important to avoid 3' end complementarity which would lead to the formation of primer-dimers (Rolfs et al,1992; Clontech Labs. Inc., 1991). The majority of primers were designed with a G or C residue at the 3' end. Primer sequences were checked for specificity against other ovine CD1 sequences using the GAP programme of UWGCG version 7.3 (Devereux et al, 1984).

Various precautions were taken to avoid contamination of the PCR reactions, including

Table 1: PCR Reaction Details and Cycling Parameters.

The reaction buffers, template DNA, primers and cycling parameters used in the various PCR amplifications carried out during this project are summarised in the table. Details of the primer sequences are given in Appendix 2.

1. The PCR cycling parameters show the denaturation, annealing and extension temperatures, as well the time taken (in minutes) for these phases during each PCR cycle. The single temperature of 72°C quoted in several of the PCR's is the final extension temperature, generally held for 10 minutes.

2. PCR amplification of the human CD1C probe required the use of sterile toothpicks to "stab" M13 plaques. The plaques were then deposited in the PCR solution by rolling the toothpicks between thumb and first finger with the plaque end immersed in the PCR mix.

3. Amplification of the 5' ends of the A25 and B-42 sequences required nested PCR reactions. The primers used in the first round of PCR are given after the 1. and primers used in the nested reaction are given after the 2. Two different PCR approaches were employed for 5' end amplification - (**A**) the RACE-PCR protocol and (**B**) the PCR protocol involving the 5' degenerate primer B199.

	human CD1 α3 probe PCR	Ovine A25 α3 PCR	ovine T10 α3 PCR	5'ends PCR for A25 and B-42	ovine gut CD1 PCR	ovine ATPase control PCR
PCR buffer	basic buffer	Ohara's buffer	basic buffer	Ohara's buffer	basic buffer	basic buffer
PCR buffer additions	5% DMSO	_	_	_	5% w1	5% w1
PCR Cycling Parameters	95°C-0.6' 50°C-1.0' 73°C-2.5' 30 cycles	94°C-0.6' 54°C-0.6' 72°C-1.5' 30 cycles	94°C-0.7' 55°C-1.0' 72°C-1.5' 35 cycles 72°C-10'	95°C-5.0' 55°C-3.0' 72°C-4.0' 1 cycle 94°C-0.7' 55°C-1.0' 72°C-2.0' 35 cycles 72°C-10'	hotstart 94°C-3.0' 55°C-2.0' 72°C-3.0' 1 cycle 94°C-1.0' 55°C-1.2' 72°C-2.0' 35 cycles 72°C-10'	hotstart 94°C-0.7' 55°C-1.0' 72°C-1.5' 35 cycles 72°C-10'
template DNA	M13 plaque stab (see 3.1.2)	1μl of 1/100 dilution from dsDNA miniprep. of clone A25	3μl of 1/80 dilution from dsDNA miniprep. of clone T10	10-15μl cDNA from 400μl cDNA pool (see 2.13)	3μl cDNA from 40μl cDNA pool (see 2.13)	3μl cDNA from 40μl cDNA pool (see 2.13)
primers	M13 for and rev seq. primers	5801 5802	G5935 G5936	A25: 1.B199+C334 2.B199+B202 B-42: 1.B199+C333 2.B199+B203	B199 G7206 (same primers for nested PCR)	ATPase 1 ATPase 2
nested PCR reaction?	no	no	no	yes- template was 1μl of first round product	yes- template was 3μl of first round product	no

wearing gloves at all times and changing them frequently, assembling reaction components in a laminar flow hood or setting up PCR reactions in a laboratory where no CD1 cloning had previously been carried out, and using positive displacement pipettes to reduce aerosol contamination.

3.1 Library Screening and Identification of Ovine CD1 cDNA Clones.

3.1.1 Introduction

The initial objective of the work outlined in this thesis was to identify and sequence complementary DNA encoding the individual ovine CD1 molecules. Previous molecular studies of CD1 had been carried out in humans, mice and rabbits, with broadly similar strategies employed in each case. The first human CD1 cDNA clone was identified by isolating a cell line (NH17) expressing high levels of HTA-1 (CD1A molecule), enrichment of HTA-1 mRNA by immunoselection of polysomes, synthesis of cDNA, and screening of cDNA clones by mRNA selection followed by *in vitro* translation and immunoprecipitation (Calabi and Milstein, 1986). One clone (FCB6) was found to select a mRNA encoding an HTA-1 α chain like polypeptide according to relative molecular mass and recognition by a specific rabbit antiserum. Clone FCB1 was subsequently identified from a total NH17 cDNA library using FCB6 as a probe. Each clone encodes the HCD1A molecule. There are three single base differences between the two clones which are probably cloning artefacts, with FCB1 extending 180 nucleotides further in the 5' direction.The FCB1 clone begins approximately 30 nucleotides upstream from the start of the $\alpha 2$ domain. A Southern blot of human Molt4 DNA digested with EcoR I and probed with the insert from FCB6 identified five hybridising bands (Martin et al, 1986). These were found to correspond by size to the EcoR I fragments of five different clones isolated from a genomic library (R4B3, R1L5, R7L4, R3G1 and R2G4), and represent the five different human CD1 genes (HCD1A, B, C, D and E). The $\alpha 3$ domains of the genomic clones were sequenced and the FCB6 cDNA clone was matched with the λR4B3/CD1A clone. In 1989, Balk et al reported the isolation of a cDNA clone corresponding to HCD1D (Balk et al, 1989a). This clone was isolated by screening a thymus cDNA library with a mixture of full length cDNA CD1A,B and C probes. They used the same probes, followed by probes corresponding to the CD1D 5'UT, $\alpha 1$ and 3'UT domains, to identify the equivalent genomic clone.

In the mouse, the first CD1 gene (MCD1.1) was identified by screening a genomic library with

the insert from FCB6 (Bradbury et al,1988). The second gene was identified by screening a genomic library with a subclone of MCD1.1 which spans the $\alpha3$ region, to yield MCD1.2. The insert from clone FCB1 was used to screen a thymus cDNA library and nine positive clones were identified (Bradbury et al,1990). None of these clones is correctly spliced, as discussed in 3.1.5. Balk et al (1991a) also identified several identical murine CD1 cDNA clones, by screening a thymus cDNA library with a probe derived from the gene encoding the $\alpha2$ and $\alpha3$ domains of HCD1D. Partial sequencing showed that this CD1 is identical to the MCD1.1 previously identified by Bradbury et al. The mouse CD1 genes are most homologous to HCD1D. The sequence of another rodent CD1 gene, rat CD1, has recently been reported (Ichimiya et al, 1994). The cDNA encoding rat CD1 was isolated from a thymic lymphoma cDNA library using a PCR-derived rat $\alpha3$ region probe. The rat CD1 sequence is highly homologous to both murine sequences in the 5'UT region and in the coding domains. Sequence divergence in the 3'UT region indicates that rat CD1 is equivalent to the murine CD1.1 gene (80% identity) and not to the CD1.2 gene (52% identity). In the rabbit, screening of cottontail and domestic rabbit genomic libraries with the insert from FCB1 and rescreening of positive clones with a synthetic oligonucleotide mixture corresponding to all sequences which could possibly encode residues 10-15 of the known rabbit CD1 polypeptide sequences, identified one clone from each library (Calabi et al, 1989b). The cottontail rabbit clone encodes the $\alpha1$, $\alpha2$ and $\alpha3$ domains and is most related to the human and mouse CD1D genes.The domestic rabbit clone encodes the 5'UT and $\alpha1$ domains only, and is most similar to HCD1B.

All of the known CD1 sequences except for clone FCB6 have been identified by library screening techniques. An alternative approach for identifying the individual ovine CD1 cDNA's would be to use the potentially faster PCR technique to amplify the required sequences (Mullis and Faloona,1987). However, in practice, the use of PCR can itself prove rather laborious, due to the necessity of sequencing multiple clones derived from a single PCR product. The sequencing of multiple clones is required to produce a consensus sequence because of the relatively high error rate of the Taq polymerase enzyme (Eckert and Kunkel, 1991; Tindell and Kunkell, 1988). PCR also requires areas of well conserved sequence across species barriers in order to generate appropriate primers for the PCR. If

these regions lie within the coding sequence this generates the additional problem of obtaining sequence information for the 3' and 5' ends.

The strategy I chose to follow for identification of the ovine CD1 DNA sequences was to screen a variety of sheep thymocyte cDNA libraries with heterologous human probes. Previous monoclonal antibody studies of sheep tissues had confirmed the expression of CD1 molecules in the thymus cortex (Dutia and Hopkins,1991), and several thymocyte libraries were provided for screening by Dr. Wayne Hein, Basel. A foetal liver library was also screened, since several reports had provided evidence for expression of various human and murine non-classical class I and CD1 molecules in the liver, particularly during foetal development (Houlihan et al,1992; Balk et al,1991a; Bleicher et al,1990; Bradbury et al, 1988). The foetal liver cDNA library screened was derived from lambs at day 40-42 of gestation. It is around this time that the first identifiable lymphocytes are found in the thymus of the foetal sheep. Liver CD1 expression has also recently been observed in the sheep (S.Rhind, personal communication).

3.1.2 Libraries, Probes and Preparation For Use.

Two human CD1 probes were a kind gift of Dr. Cesar Milstein, MRC Laboratory of Molecular Biology, Cambridge, one a CD1A probe and the other a CD1C probe. The CD1A genomic clone B353.88 contains a ~0.6 kb insert comprising the coding sequence for the human CD1A exon 4/α3 domain (0.28kb) as well as some flanking intron sequence (Martin et al, 1986). This had been prepared from a random shotgun library and blunt end cloned into the Sma I site of M13tg130. The CD1C genomic clone L4R.J4 contains a ~0.6 kb insert comprising the coding region of the human CD1C exon 4/α3 domain (0.28kb) as well as some flanking intron sequence (Martin et al,1986). The insert had been isolated and cloned in an identical manner to the CD1A clone.

Four different sheep cDNA libraries were screened during this project- a foetal thymocyte library (termed SFT for sheep foetal thymocyte; Hein et al,1989), two lamb thymocyte libraries (termed ST-1 and ST-2; Hein et al, 1989; Grossberger et al,1990), and a foetal liver library

(termed SFL). The thymocyte libraries were generated by isolation of RNA from the thymus of two single eight week old lambs (ST-1 and ST-2) and from the pooled thymi of three 67day old foetal lambs (SFT). The SFL library was generated from the pooled livers of twelve 40-42 day old foetal lambs (SFL). Poly (A) $^+$ RNA was isolated, cDNA synthesised and a poly C tail added. Oligo (dG) was used to prime second strand synthesis, and the dsDNA cloned into the Eco RI site of the vector λgt11. All libraries were a kind gift from Dr. Wayne Hein, Basel Institute for Immunology, Basel, Switzerland.

In order to use the human probes for screening the cDNA libraries, double-stranded DNA (dsDNA) from the M13 human clones was required. The initial approach taken was an attempt to isolate the double stranded replicative form (RF) of M13 from which insert DNA could be isolated. However, despite several attempts, isolation of the RF form was never achieved, although it was relatively easy to isolate single stranded 'phage DNA from the culture supernatants. An alternative approach was then taken. This involved the use of the PCR reaction and M13 forward and reverse sequencing primers (see Appendix 2) to amplify insert DNA taken from "plaque stabs" (DNA Cloning vol.1).

A titration of the provided M13 human CD1C clone DNA was performed in order to determine the dilution of M13 DNA which would provide well-separated plaques to stab for PCR. Essentially, 50μl PCR solutions were set up as described in Methods, 2.14. Figure 1A illustrates the result of a set of plaque stab PCR's, showing that fragments of the expected size were obtained. The PCR fragments were subsequently excised from the gel, pooled, and purified by phenol/chloroform and chloroform extractions and sodium acetate precipitation. The insert DNA was prepared for cloning using the Klenow fragment of *E. coli* DNA polymerase I to blunt end the DNA and T4 polynucleotide kinase to phosphorylate the 5' ends, and cloned into the vector pTZ19R (FIG. 1B). A large scale plasmid preparation of dsDNA from sample number 6 was carried out as described in 2.10.2. This yielded ~3.5mg of plasmid DNA from a 1 litre culture. The identity of the insert as human CD1C was confirmed by sequence analysis of the recombinant dsDNA with the M13 forward and reverse primers. A sample of the recombinant dsDNA was digested with the enzymes EcoR I and Hind III, yielding

Figure 1: Preparation of CD1 Probes From M13 Human Genomic CD1 Clones.

Photographs A, B and C illustrate the results of PCR amplification of the HCD1C insert from the L4R.J4 M13 clone analysed by gel electrophoresis.

A. Results from seven plaque stab PCR's, as outlined in the text. 25% of each PCR reaction was analysed by gel electrophoresis, and five of the seven samples contained inserts of the expected size (0.6kb).

B. The remaining PCR mixes from the five positive samples shown in A were purified by phenol/chloroform extraction, blunt ended with Klenow and phosphorylated with kinase, gel purified, and the inserts ligated into Sma I cut pTZ19R. ssDNA was prepared from six samples and 3μl from a total of 30μl were digested with EcoR I and Hind III. The results from three samples are shown, with insert bands of 600bp apparent in lanes 2, 4 and 6. Lanes 1,3 and 5 show equivalent amounts of undigested samples.

C. A largescale dsDNA plasmid preparation from sample six (B, lane 4) was carried out. 30μg of DNA was digested with EcoR I and Hind III and separated on a gel over four lanes (2-5). Lane 1 shows undigested pTZ plasmid. Results of the digest show bands of the expected size (0.6kb). The four bands were pooled and purified prior to use as a probe in library screening.

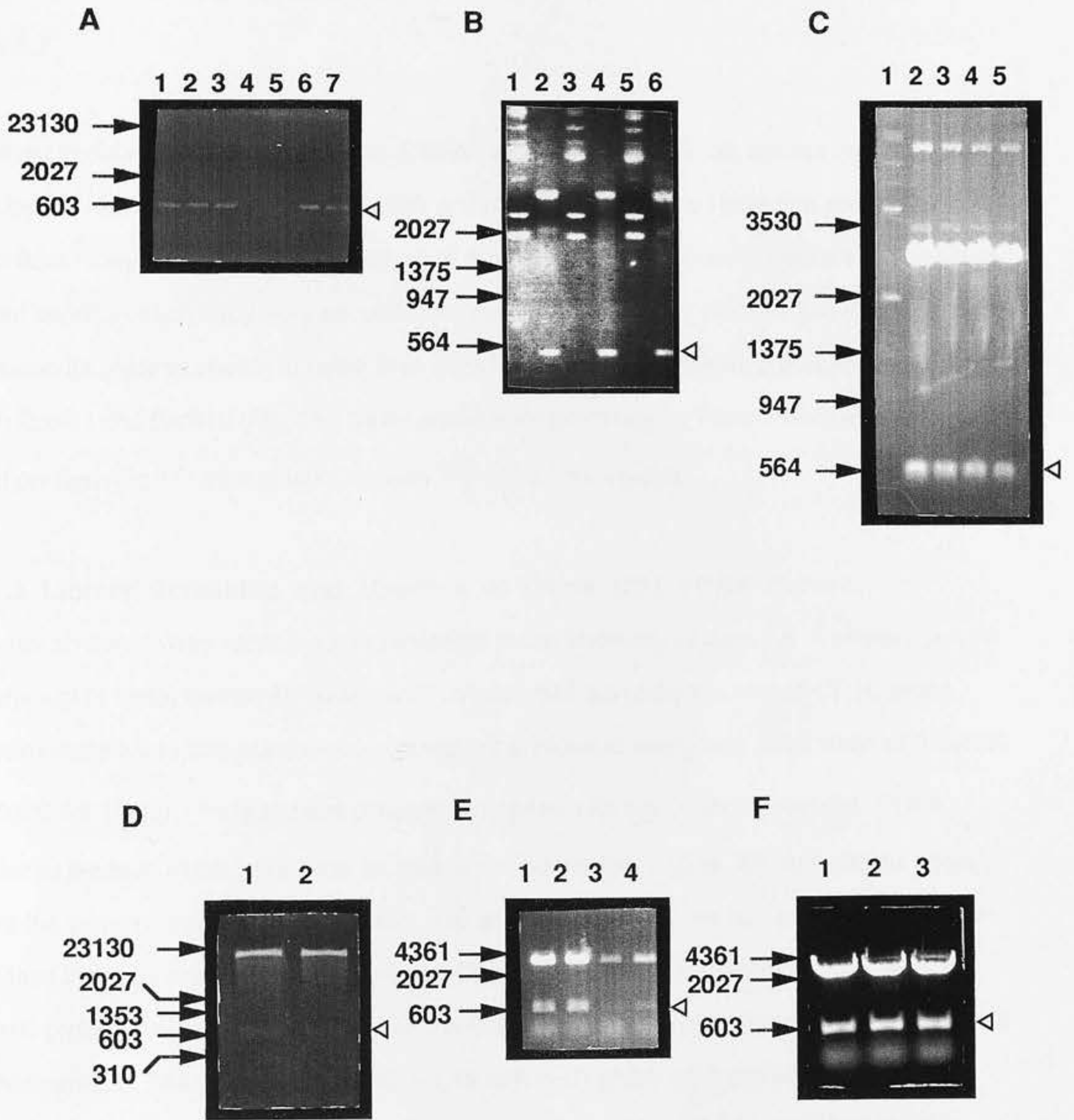
Photographs D, E and F illustrate preparation of the HCD1A probe analysed by gel electrophoresis.

D. 15μl of a 1/10 dilution of the provided M13 CD1A clone DNA (B3S3.88) was digested in a total volume of 30μl with Bam HI and EcoR I, and separated by electrophoresis over two lanes, yielding an insert of the expected size (0.6kb).

E. The two faint insert bands from 1D were excised, purified and ligated into EcoR I /Bam HI digested pBluescript. Four white colonies of transformed *E.coli* were picked for DNA isolation, 1μl from a total of 50μl of the dsDNA was digested with EcoR I and Bam HI and analysed by gel electrophoresis. All four clones yielded insert bands of the expected size (0.6kb), although there was much less DNA in samples 3 and 4.

F. A further Bam HI/EcoR I digest was carried out on 20μl from the 50μl total of dsDNA from sample 1 and analysed by electrophoresis over three lanes. This provided sufficient insert DNA for purification and use as a probe for library screening.

Molecular weight markers are indicated in base pairs.


A
1 2 3 4 5 6 7
23130
2027
603
B
1 2 3 4 5 6
2027
1375
947
564
C
1 2 3 4 5
3530
2027
1375
947
564
D
1 2
23130
2027
1353
603
310
E
1 2 3 4
4361
2027
603
F
1 2 3
4361
2027
603

sufficient insert for purification and use as a probe (FIG. 1C). The four insert bands were excised from the TAE gel, pooled, purified and labelled with ^{32}P as described in sections 2.6 and 2.7.

In order to obtain dsDNA of the human CD1A fragment as a probe, an attempt was made to excise the insert band from the M13 DNA and clone it directly. The restriction enzymes EcoR I and Bam HI were used to digest a sample of the M13 DNA, as shown in Figure 1D. The single insert bands in each track were excised from the TAE gel, pooled, purified and ligated into the pBluescript plasmid vector. Plasmid DNA from four white colonies was prepared, and digested with EcoR I and Bam HI (Fig. 1E). Insert bands from preparations 1 and 2 were excised, pooled and purified (Fig.1F) prior to labelling with ^{32}P for use as a probe.

3.1.3 Library Screening and Isolation of Ovine CD1 cDNA Clones.

Library screening was carried out as described in the Methods, section 2.8. A primary screen of the λgt11 foetal thymocyte library (SFT) was carried out using the human CD1C probe. Approximately 240,000 plaques were screened at reduced stringency (final wash of 0.5xSSC at 58°C for 1 hour). Forty positive plaques were picked for secondary screening. These included plaques which gave weak as well as strong positive signals. Using duplicate filters from the same primary plates, this library was also screened with the human CD1A probe, as detailed in a later section of 3.1.3. Positive plaques from the SFT library screened with the CD1C probe were called series A, while positive plaques identified with the CD1A probe were called series B. The primary screening results with each probe were directly compared by aligning the filters over a light box. Very few plaques were identified by both probes. This result was rather unexpected since it was presumed that the CD1A probe would identify most if not all of the ovine CD1 clones, generating a major overlap with plaques identified by the CD1C probe.

Lambda DNA (see 2.8.4) was prepared from the positive plaques identified in the secondary screens, followed by digestion of the DNA to determine if any EcoR I excisable inserts were present. Details of the results of screening the various libraries are given below.

Screening the SFT Library With HCD1C (series A).

Forty positive plaques were picked for secondary screening, of which thirty-one were initially plated at a suitable density for secondary screening. Seventeen of the thirty-one samples in the secondary screen gave positive signals, and lambda DNA was prepared from all of these. Four samples yielded EcoR I excisable inserts- A9, A16, A25 and A31. The insert band sizes were approximately 750, 950, 1000, and 1100 nucleotides respectively. The insert bands were excised from the TAE gel, purified, and ligated into the EcoR I-cut pBluescript plasmid vector. Transformed *E.coli* were plated out on the appropriate selective media and colonies picked for DNA isolation on the basis of blue/white colour selection (2.2.1).

Double stranded plasmid DNA was prepared from four transformants derived from each sample, followed by EcoR I restriction enzyme digests to check the inserts, and the cleanest preparations chosen for sequencing (2.10). Of the four secondary screen positive clones, A16, A9 and A31 were found to be negative for CD1 by initial sequencing with the primers 392 and 393. However, the A25 clone (insert size ~1000 bp's)was identified as an ovine CD1 by comparison of the initial sequence data with the known human sequences. This comparison showed A25 to be most homologous to human CD1B. Figure 2 illustrates the A25 insert excised from the lambda DNA (2A) and excised from the pBluescript plasmid after subcloning (2B).

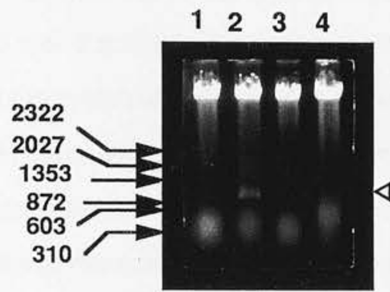
At this point it seemed appropriate to carry out further secondary screens on three samples which had been plated too densely for secondary screening using the initial dilution of the 'phage stock, as well as five other samples which had not given positive hybridisation signals on the initial secondary screening. All eight samples gave positive signals this time, and preparation of lambda DNA followed by EcoR I digestion yielded inserts from two further samples, A23 and A12, with insert sizes of approximately 1000 and 2000 nucleotides. The inserts were isolated and subcloned into pBluescript, so that dsDNA could be generated for sequencing analysis. The initial sequencing results indicated that A12 is not homologous to the human CD1 sequences, and that A23 is identical to the previously cloned A25 ovine CD1.

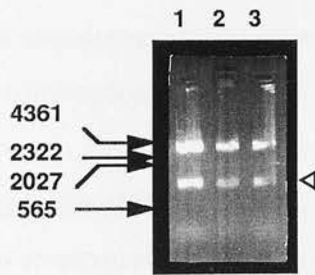
Figure 2: Isolation and Subcloning of SCD1A25

Panel A shows Eco RI digests of lambda DNA from four positive clones analysed by gel electrophoresis. The clones were isolated from the SFT library by screening with the human CD1C probe. The ~1000bp insert excised from clone A25 is visible in lane 2, flanked by three Eco RI insert negative clones.

Panel B shows Eco RI digests from three transformants analysed by gel electrophoresis after subcloning of the A25 insert into the pBluescript plasmid.

Molecular weight markers are indicated in base pairs.


A
1 2 3 4
2322
2027
1353
872
603
310


B
1 2 3
4361
2322
2027
565

Since thirteen of the seventeen clones which gave positive signals after secondary screening did not yield EcoR I excisable inserts, a different approach was applied. The enzyme Mlu I cuts at several points in the ~~pBluescript plasmid~~ lambda vector, including at either side of the polylinker site, so that approximately two kilobases of ~~pBluescript~~ lambda vector DNA flanks any insert at this site. Therefore, an Mlu I digest of ~~pBluescript~~ lambda vector which contains an insert would yield, among others, a band of x+2kb, where x is the actual size of the insert in kb. Thirteen samples were digested with Mlu I, and seven samples yielded "insert" band sizes of 3000-3200 bases. These were excised, purified, blunt ended (see 2.3.2) and subcloned into the EcoR V site of pBluescript. Since the EcoR V site is not reconstituted after ligation of the insert DNA into the plasmid, other enzymes which cut in the polylinker site on either side of the insert were used in order to excise it from the plasmid DNA. The enzymes Sac II and Hinc II were chosen as they both work equally well in a particular restriction enzyme buffer. However, digests of ds plasmid DNA with these enzymes gave some unexpected results. It was expected that either single bands of approximately 3 kb in size, or multiple insert bands totalling approximately 3 kb, would be observed following digestion and gel electrophoresis, depending on the existence of internal Sac II or Hinc II sites within the insert. In fact, a variety of insert sizes were observed, few of which corresponded to the expected results. Four clones derived from different samples were checked by dsDNA sequencing with the primers 392 and 393, and initial results indicated that none are homologous to the human CD1 sequences.

Screening The SFT Library with HCD1A (series B).

The SFT library was also screened with the human CD1A probe to give series B samples. Again, approximately 240,000 plaques were screened at reduced stringency (final wash of 0.5xSSC, at 58°C for 1 hour). Forty positive plaques were picked, of which thirty-two samples were plated at a suitable density for secondary screening using the initial dilution of the 'phage stock. Of the thirty-two samples in the secondary screen, eighteen produced positive plaques. Twelve samples yielded EcoR I excisable inserts, although the results were a little unexpected. Two of the samples contained double inserts of approximately 1 and 3 kb in each case. For each sample, both bands were isolated for subcloning (B8U and L, B34U and L). Eight of the other ten insert-positive samples contained inserts of approximately 3000

bases. The remaining two samples, B16 and B21, contained inserts of approximately 4000 and 6000 bases respectively. These inserts seemed too large given that the expected full length coding region size of CD1 molecules is approximately 1 kb, although mRNA transcript sizes ranging from 1.2 to a maximum of 3kb have been observed in humans and mice (Calabi et al, 1988; Calabi et al, 1989c). Inserts from B11 and B21 were isolated and purified for subcloning, in addition to the B8 and B34 inserts. Sequencing reactions were carried out on plasmid samples from B8L, B11, B34U and B34L, and initial sequencing results indicated that none of the samples are homologous to human CD1 sequence.

Due to the problems with Mlu I digests and subsequent cloning of inserts from the series A samples, it was decided not to pursue this approach for the B series. However, it was possible to carry out a further secondary screen on five samples which had previously been plated too densely for rescreening, and on five samples which had previously been negative on secondary screening. A further seven EcoR I excisable inserts were identified. Of these, three samples contained inserts of 4000 bases or more (B2, B6 and B22), and four contained inserts of approximately 3000 bases (B30, B4, B13 and B36). Again, these insert sizes were unexpectedly large considering the coding region size of CD1 molecules. Initial sequence analysis of selected clones from B30,B4, B13 and B22 indicated that none are homologous to human CD1.

The A25 clone (renamed SCD1A25 for subsequent use- for sheep CD1, A25 clone) had been completely sequenced by this point, allowing a homologous sheep probe to be derived by PCR amplification of the $\alpha3$/TM/CYT domain for use in further library screening (Figure 3A). A 1 μl sample of a 1/100 dilution of the SCD1A25 ds clone DNA was used as a template for the primers 5801 and 5802, with cycling carried out as described in Table 1 (section 2.14). Gel electrophoresis of the resultant PCR product showed that a band of the expected size (~450 nucleotides) had been amplified (Fig.3B, Lane 1), as well as two less prominent higher molecular weight bands. The 450 bp band was purified using a Spinbind column (Fig.3B, lane 2). However, as the two upper bands were as efficiently recovered as the required band, the remainder of the purified sample was run on a TAE gel, and the main band excised from the

Figure 3: Preparation of the PCR Derived A25 α3/TM/CYT Region Probe.

Panel A shows a partial sequence of the insert from the SCD1A25 cDNA clone, with the borders of the α3 region indicated by vertical dashed lines. The stop codon is denoted by a line under the trimer sequence TGA. The primers 5801 (derived from the sense strand) and 5802 (derived from the antisense strand), as indicated in bold type, were used to amplify the α3/TM/CYT region by PCR, using 1μl of a 1/100 dilution of dsDNA from the SCD1A25 clone as a template.

Panel B shows a photograph of the PCR product analysed by gel electrophoresis.

Lane 1: 2μl of PCR product from a total of 30μl amplified.
Lane 2: The remaining 28μl of PCR product was purified on a Spinbind column, eluted in 20μl and 1μl analysed by gel electrophoresis.

Molecular weight markers are indicated in base pairs.

A

5801

239 ATATTT**CCTG AGTGTCCTTG ATGC**AGGCAA AGCAGAACTG CAGAGGCAAG

|Alpha 3 domain

289 |TGAAGCCTGA GGCCTGGCTG TCCAGCGGCC CCACTCCTGG GCCTGGCCGC

339 CTACTGCTGG TGTGCCATGT CTCAGGATTC TACCCAAAAC CTGTGTGGGT

389 GATGTGGATG AGGGGTGAGC AGGAAGAGCC TGGCACTCAG CAAGGGGACA

439 TCATGCCCAA TGCAAATTGG ACTTGGCATC TGCGAGCAAC CCTAGATGTG

489 GCAGCTGGAG AGGCGGCTGG TCTGAGTTGC CGAGTGAAGC ACAGCAGTCT

|TM/CYT domain

539 AGGAGACCAG GACATCGTCC TGTACTGGG|G ACACCCCACC TCCACTGGCC

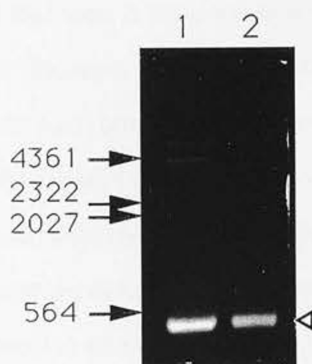
589 TGATATTTGT GGCAATAATA GTGTCCTCCC TCATCCTTTT GATCTGTCTT

639 GCATTATGGT TTTGGAGGCG CTGGTCATAT CTGACTATCT TG<u>TGA</u>**GCCCA**

689 TCACTGTCTC CTTTTCCATT

5802

B



gel prior to further purification using a Spinbind column (see section 2.6). This removed the two higher molecular weight bands, eliminating the possibility of them interfering with library screening.

Screening The ST-1 and ST-2 Libraries With The Homologous Probe (series A- and B-).

The homologous sheep probe was subsequently used to screen two lamb thymocyte libraries, ST-1 and ST-2. Approximately 120,000 plaques were screened from each library, using stringency conditions of 0.2xSSC at 65°C for the final wash. Positive plaques from the ST-1 library were called series A-, while positive plaques identified from the ST-2 library were called series B-.

Nineteen positive plaques were identified from the ST-1 library after primary screening, of which fourteen were plated at a suitable density for secondary screening. All fourteen samples gave positive signals in the secondary screen, and lambda DNA was prepared from each. Three EcoR I excisable inserts were identified, in samples A-14, A-64 and A-81, with insert sizes of approximately 1000, 2000 and 2000 bases respectively. For the ST-2 library, six positive plaques were identified in the primary screen, of which four were initially plated at an appropriate density for rescreening. All gave positive signals after the rescreen, and lambda DNA was prepared from each one. Two EcoR I excisable inserts were identified, in samples B-42 and B-52, with insert sizes of approximately 1000 and 1200 bases respectively. Inserts from the five positive samples (A-14, A-64, A-81, B-42, B-52)were subcloned into the pBluescript plasmid, and dsDNA prepared for sequencing. Initial sequencing showed that two of the inserts are identical in size and sequence (A-14 and B-52), and are ovine CD1 equivalents. B-52 was arbitrarily retained for further study and renamed SCD1B-52. B-42 was also retained for further sequencing, and was renamed SCD1B-42. Figure 4A shows the B-42 and B-52inserts excised from the lambda DNA and Figure 4B shows the same inserts excised from the pBluescript plasmid after subcloning.

Initial sequencing of clones A-64 and A-81 revealed poly A tracts at both ends of each clone. It thus appeared that two inserts had originally been cloned in a head to head orientation into

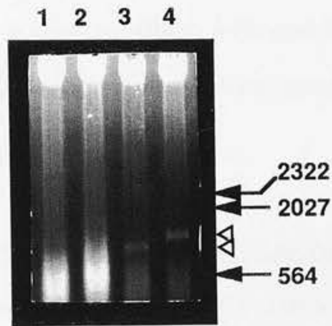
Figure 4: Isolation and Subcloning of SCD1B-42 and SCD1B-52.

Panel A shows Eco RI digests of lambda DNA from four positive clones analysed by gel electrophoresis, with B-42 in lane 3 (~950 base pairs) and B-52 in lane 4 (~1100 base pairs).The samples analysed in lanes 1 and 2 did not yield Eco RI excisable inserts.The clones were isolated from the ST-2 library by screening with the homologous α3/TM/CYT region probe.

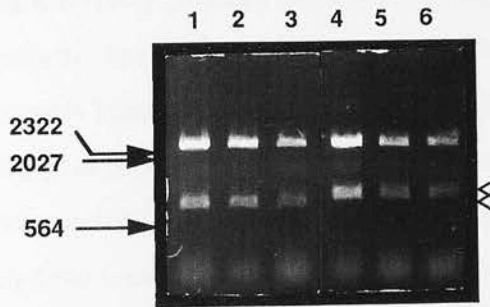
Panel B shows Eco RI digests from three transformants each of B-42 (lanes 1-3) and B-52 (lanes 4-6) analysed by gel electrophoresis after subcloning the lambda inserts into the pBluescript plasmid.

Molecular weight markers are indicated in base pairs.

A



B



the EcoR I site of the lambda DNA when the cDNA library was generated. The EcoR I site between the two inserts appears to have been lost, since only a single insert of approximately 2000 nucleotides was isolated from each lambda clone. The sequence from A-64 and A-81 also revealed that one of the two presumed inserts in each clone is virtually identical to the 3' end of the B-42 clone. The other inserts in A-64 and A-81 are identical but show no homology at all to the known CD1 sequences. No further sequence analysis was pursued for these two clones.

Screening The SFT and SFL Libraries With The Homologous Probe SCD1A25 α3/TM/CYT.
The PCR generated homologous α3/TM/CYT probe was used to screen the SFL library and rescreen the SFT library. Approximately 150,000 plaques were screened from each library. The hybridised filters were washed with 0.5xSSC at 65°C for forty minutes, with a final stringent wash of five to ten minutes in 0.2xSSC at 65°C. No positive plaques were identified from the foetal liver library, although this may have been due partly to the stringency of the washing conditions. Twenty positive plaques were isolated from the foetal thymocyte library. Of these, four were initially plated too densely for secondary screening. Of the remaining sixteen, twelve gave positive signals after secondary screening. Lambda DNA was prepared from all of these and enzyme digestion yielded one EcoR I excisable insert (T10) of approximately 1200 nucleotides (Fig. 5A). The insert was excised from the TAE gel, purified, and ligated into the pBluescript plasmid for sequencing (Fig. 5B). Initial sequence analysis confirmed that the insert was homologous to human CD1. The sample was renamed SCD1T10 for further analysis.

3.1.4 Presentation of cDNA Clone Sequence Data.

The four ovine cDNA clones, SCD1A25, SCD1B-42, SCD1B-52 and SCD1T10, were fully sequenced on each strand by dsDNA sequencing using the dideoxy chain termination technique (Sanger et al,1977). The complete sequence of each cloned insert was determined by primer walking, rather than subcloning. Primer walking involves the synthesis of an oligonucleotide which hybridises near the 3' end of the newly obtained sequence, and which primes synthesis in a subsequent set of dideoxy reactions (Strauss et el, 1986). The

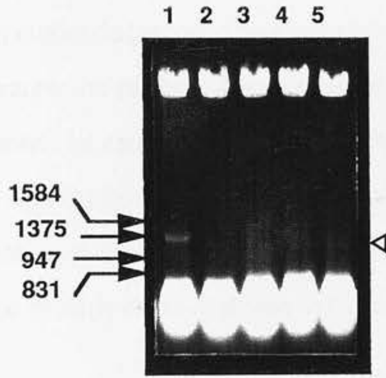
Figure 5: Isolation and Subcloning of SCD1T10.

Panel A shows Eco RI digests from five positive clones analysed by gel electrophoresis. The clones were isolated from the SFT library by screening with the homologous α3/TM/CYT probe. The ~1400 base pair insert from T10 is visible in lane 1, followed by four clones from which no Eco RI inserts were obtained (lanes 2-5).

Panel B shows Eco RI digests from three transformants analysed by gel electrophoresis after subcloning of the T10 insert into the pBluescript plasmid.

Molecular weight markers are indicated in base pairs.

A



B

sites chosen for the sequencing primers were located at 80-100 nucleotides upstream from the 3' end of the most recently obtained sequence in order to generate some overlap between sequences derived from different primers. The primer sequences chosen were generally 18-20 nucleotides in length and care was taken to avoid the potential for hairpin loop formation within each primer. The sequences obtained for the sense and antisense strands of each clone were compared and any ambiguities resolved by referring back to the original sequencing autoradiographs. The sequencing strategy employed for each of the cDNA clones, illustrating the position of the sequencing primers, is presented in Figure 7. The complete sequence of each primer is given in Appendix 2. The sequences generated were assembled using the University of Wisconsin Genetics Computer Group (UWGCG) version 7.3 sequence analysis software package (Devereux et al, 1984). The complete nucleotide sequence of each clone is shown in Figure 6, A-D.

A summary of the domain structure of each clone, as well as the sequencing strategy used, is outlined in Figure 7. Comparison of the sequence of each of the four ovine CD1 clones to the known human, mouse and rabbit sequences revealed that all were most homologous to the human CD1B gene (a detailed analysis of these comparisons is provided in section 3.3). The positions of the domain boundaries in the ovine cDNA clones were therefore assigned by comparison to the known genomic exon/intron boundaries of the human CD1B gene (Martin et al, 1987; Aruffo and Seed,1989), as exon/intron boundaries tend to be highly conserved across species barriers.

As outlined in Figure 7, both SCD1A25 and SCD1B-42 were isolated as truncated versions of the CD1 sequence. The SCD1A25 sequence begins towards the end of the $\alpha1$ region, and the SCD1B-42 sequence begins fifty-eight nucleotides into the $\alpha2$ region. Section 3.2 outlines the approach taken to determine the sequence of the 5' untranslated (5'UT) and $\alpha1$ regions of these two clones. The SCD1T10 clone was also isolated as a truncated form, with the sequence beginning towards the end of the $\alpha1$ domain. However, attempts to isolate the 5' region of this CD1 subtype were unsuccessful (see section 3.2). The SCD1B-52 clone is the only clone which contains the 5'UT and $\alpha1$ regions, but it lacks an $\alpha3$ domain. This

Figure 6: Sequences of The Four Ovine CD1 cDNA Clones.

The four sequences illustrated in A-D represent the sequence information obtained by dsDNA sequencing of each of the four ovine cDNA clones. Each clone was fully sequenced on each strand by primer walking (see Fig.7).

Fig.6A SCD1A25
Fig.6B SCD1B-42
Fig.6C SCD1B-52
Fig.6D SCD1T10

The domain boundaries of each clone are indicated by vertical dashed lines.

Long sequences (~20 nucleotides) marked in bold type represent primers derived from the sense strand, while long sequences which are underlined denote primers derived from the antisense strand.

Trimer sequences marked in bold type represent either the start codon (ATG) or the stop codon (TGA).

The six-mer sequences marked in bold type in the SCD1A25, SCD1B-42 and SCD1B-52 sequences denote the location of the polyadenylation signal in each sequence.

Fig.6A: SCD1A25

```
              Alpha 1 Domain   |Alpha 2 Domain
1                 CAGCTGGA AT|ACCCCTTT GTGATCCAGG ACATAGCAGG

39   TTGTGAGCTG CATCCTGGGA AGGCCGTAGA AAGCTTCTTG AAGGGAGCTT
                          G5784
89   TTGGAGGATT GGATTTCGTG AGCATCAAGA ATGATTCATG TGCACCTGTC

139  CCAGGAGGCG GCAGCATGGC CCAGCGTTTC TATGAACTCA TCATTCAGAC

189  CATGCTATCT GTGATACTAT AGCTAAGCTC CTCTTAGAAA CCTGCCCTCG
                     5801
239  ATATTTCCTG AGTGTCCTTG ATGCAGGCAA AGCAGAACTG CAGAGGCAAG
     |Alpha 3 Domain
289  |TGAAGCCTGA GGCCTGGCTG TCCAGCGGCC CCACTCCTGG GCCTGGCCGC

339  CTACTGCTGG TGTGCCATGT CTCAGGATTC TACCCAAAAC CTGTGTGGGT

389  GATGTGGATG AGGGGTGAGC AGGAAGAGCC TGGCACTCAG CAAGGGGACA
                                B200
439  TCATGCCCAA TGCAAATTGG ACTTGGCATC TGCGAGCAAC CCTAGATGTG
                                                 B201
489  GCAGCTGGAG AGGCGGCTGG TCTGAGTTGC CGAGTGAAGC ACAGCAGTCT
                                      |TM/CYT Domain
539  AGGAGACCAG GACATCGTCC TGTACTGGG|G ACACCCCACC TCCACTGGCC

589  TGATATTTGT GGCAATAATA GTGTCCTCCC TCATCCTTTT GATCTGTCTT

639  GCATTATGGT TTTGGAGGCG CTGGTCATAT CTGACTATCT TGTGAGCCCA
        5802
689  TCACTGTCTC CTTTTCCATT TGGAATAAGT ACCCACGAAC CCAGAAACTC

739  AAGCGGTCAG CCCAGTAGCC AATTTCATCA TATTTCATCA AATAATCATC

789  TGATCAAATC AAGTTCCTGT AGATTGCAAG ATCAGAAACA GAATTAAAGG

839  TGTTCATATA TTATGAGACA ATACTAATAT TAGGATCTGC TCAGATTTCT

889  TGGATGTGGG ATGCGGGAAA GAATGTACCT TGTAATAAAT GAAATGATGT

939  ACACAAAAAA AAAAAAAAAA AAAAAAA
```

Fig.6B: SCD1B-42

1 AAAGCTCT TTGAGAGGAGC

Alpha 2 Domain

20 CTTTAGGAGG ACTGGATGTT TTGAGGATCC AGAATCATTC CTGCATGCCT

70 GCACCAGACA GCGGCAACAG GGGGCAGAAG CTTTGTGCAC TCCTGAGTCA

120 GTATCAAGGC ACCTCCGATA TCATTGAGAG ACTCGTCTCA GAAACCTGTC

170 CTCGATATCT CCTGGGTGTC CTCGATGCAG GGAAGGCAGA ACTGCAGAGG

|**Alpha 3 Domain** **6088**

220 CAAG|TGAAAC **CTGAAGCCTG GCTTTCCAGT** GGCCCCACTC CTGGGCCTGG

B204

270 CCGCCTACTG CTGGTGTGCC ATGTCTCAGG ATTCTACCCA AAACCTGTGC

320 AGGTGATATG GATGAGGGGC AAGCAGGAGC AGCCTGGCAC TCAGCAAGGA

B205

370 GACATCATGC CCAATGCTGA CTGGAC**TTGG TATCTCCGAG TAAC**CCTAAA

420 TGTGGCAGCT GGGGAGGCGG CTGGCCTGAG TTGCCGAGTG AAGCACAGCA

|**TM/CYT Domain**

470 GTCTAGGAGA CCAGGACATC ATCCTCTACT GGG|GACACCC CACATCCATT

520 GGCTTGATAC TTGTGGCAAT AATAGTGCCC TCCTTGATCC TTTCGATATG

570 CCTTGCATTA TGGTTTTGGA GACGCTGGTC ATATCAGAAT ATCTTG**TGA**G

620 CCCTGACCAT GTCTCCTTTT CATTTGGAAT AAGTATCCAG GAACCTGAAA

6086

670 CTTAAGTTGT CAGCCTAGGA GTCAATCTCA TTATATTTCA TCAAATAATC

720 ATCACATTTG ATCAAATCAT TGTTCCTGTA GGTTGTAAGA TAAATCATAA

770 TTTATACTCT AGCAGAAACA TAAATGAAAA TTGTTATTAT GAGACAATAT

820 CACTAGCAGG ATCCACTCAG ATTTCATAGA TGTGATTTGT GAGAAAGATA

870 TACCTTGG***AA TAAA***TGAAAT GATGTACACA AAAAAAAAAA AAAAAAAAAA

920 AAAA

Fig.6C: SCD1B-52

```
                                               |  Leader Domain
   1  CTCCGAGAGT CAGAAGTTCT ACTTCCCAGT GAA|ATGCTGC TTCTACCACT
                                          |Alpha 1 Domain
  51  TCTGTTACTT GGAGTTATCC TCCCAGGTGG TGAC|AATGAG GATGTGTTCC

 101  AGGGGCCAAC CTCCTTCCAT CTCAAGCAGA TTTCAACCTT TGTCAACAGC

 151  ACATGGGCTC AAAATCAAGG CTCAGGCTGG TTGGATGACT TGAAGATTCA
                                G3540          6087
 201  TGGCTGGGAG AGTGACTCGG GCACTGCCAT TTTCCTGAAG CCCTGGTCCA

 251  AGGGCAACTT TACTGATGAG GAGATGACTG AGCTGGAGGA CATCTTCCGA

 301  GCCTACTTCA TTTTCTTCAC TCAGGAAGTG CAGGATCGAG TCAATGAGTT
                |Alpha 2 Domain
 351  CCAGTTAGAA T|ACCCCTTTG TGATCCAGGT CACAGCAGGC TGTGAGCTGC
                                                    A410
 401  ATTCTGGGGA GGCCATAGAA AGCTCTTTGA GAGGTGCTTT AGGAGGACTG

 451  GATGTTTGGA GGATCCAGAA TCATTCCTGT GCGCCTGCAC CAGACAGCGG

 501  TACGAGGGGG CAGAATTTTT GTGCACTCAT GACTCAGTAT CAAGGCATCT
                 A411
 551  CCGATATCCT TGAGAGACTC CTCTCAGAAA CCTGTCCTCG ATATCTGCTA
                                                |
 601  GGTGTCCTCG ATGCAGGGAA GGCAGAACTG CAGAGGCAAG| GACACCCCAC
              TM/CYT Domain
 651  ATCCATTGGC TTGATACTTG TGGCAATAAT AGTGCCCTCC TTGATCCTTT
                     G3539
 701  TGATATGCCT TGCATTATGG TTTTGGAGAC GCTGGTCACA TCGGAATATC

 751  TTGTGAGCAC TGACCGTGTC TCCTTTTCCA TTTTGGAATA AGTATCCAGG
                               6086
 801  TACCCTGAAA CTCAAGTTGT CAGCCTAGGA GTCAGTCTCA TCATATTTCA

 851  TCAAATAATC ATCACATTTG ATCAAACATT GTTCCTGTAG GTTGTAAGAT

 901  AGATCATAAT TTATACCCTA GCAGAAACAA AAATGAAAAT TGTTATTATG

 951  AGACCATATC ACTAGCAGGA TCCACTCAGA TTTCATAGAT GTGATTTGTG

1001  AGAAAGAATA TACCTTGGAA TAAATGAAAT GATGTACACA AAAAAAAAAA

1051  AAAAAAAAAA AAAAAAA
```

Fig.6D: SCD1T10

```
               Alpha 1 Domain          | Alpha 2 Domain
   1  CGGGCTGCAG GAATTCCAGT TTGAAT|ACCC ATTTGTGATT CAGGGCATAG

  51  CAGGTTGTGA GCTGCATTCT GGGAAGGCCA TACAAAGCTT CTTGAGAGCA

 101  GGTTTTGAAG GACTGGATTT CGTGAGCATC GAGAATCACT CATGTGTGCC
                                                 G3537
 151  TGAGCCAGAG GGAGGCAGTG AAGCACAGTG GTTTTGTGTT TTCATTACTC

 201  AGTACCAAGG CATCTTGGCT ATCATAGACA GGCTCCTCTC AAAAACCTGC

 251  CCCGATATC TCCTGGGTGT CCTCGATGCA GGGAAGGCGG AACTGCACAG
         | Alpha 3 Domain
 301  GCAAG|TGAAG CCTGAAGCCT GGCTGTCCAG TGGCCCCACT CCTGGGCCTG

 351  GCCGCCTATT GCTGGTCTGC CATGTCTCAG GATTCTATCC AAAACCTGTA
                             G4252
 401  CGGGTGATGT GGATGAGGGG TGAGCAGGAG CAGCCTGGTA CTCAGCAAGG

 451  AAACATCATA CTCAATGCTG ATTGGACTTG GTATCTCCGA GTAACCCTGG
                                       G3774
 501  ATGTGGCGGC TGGGGAGGCA GCTGGCCTGA GTTGTCGAGT GAAGCACAGC
                                          |TM/CYT Domain
 551  AGTCTAGGAG ACCAGGATAT CATCCTGTAC TGGG|GACACC CCATGTACAT

 601  TGGCTTGATA TTTGTGGCAA TAATAGTGCC CTCCTTGATC CTCTTGATAT

 651  GTCTTGCATT ATGGTTTTGG AGGCGCTGGT CATATCAGAC TGTTTTGTGA
        G3775
 701  TCCCTGACCA TGTCTCCTTT CCCATTTGGA GCAATTACCC AGGAGCCCAG

 751  AAACTCAGGT TGACAGCCCA GGAGTCAATT TCACTATATT TCATCAAATA
                             G4253
 801  ACCATCATAT TTGATCAAAT TAGTGTTCTG ATAAGCTTTA GATAAATCAT

 851  AATTTATATA CTGGTGGAAA AATGATAAAG TGTACATTAT AAGACAATAT

 901  CACTAGTAGA ATCCACACGG ATTTCAGGTG TGAAGTGTGA GAAAGAATGT

 951  ATCTCAGACT AAATGAAATG ATGTGCAGGA CTAAATTGTG ACATTCCTTT

1001  GGCTTCTTAT TTTATAACAT TTTTTCACT CCTACTGAAA TATCATTTGT

1051  CAAAATGAGC TAATTGTAAT TATGTCAAGA TAATTGTATT TGCAGAGATG
        G3588
1101  ACCAGCATTT TAACTTAATT TCATTGTATC TGCTTGGATG ATTTTCTATC

1151  TTGGAGAAAC TGAGCTCCTT CTTCTTCCAG GGAGCCTTCC TTGGTCCTAC

1201  AGTAGAAGTA ACCATCCCTT GACTGATGTC ATGCAGAGAT CTACTTTCCT

1251  CTTTACTCAG AACATTGCAG TTCTTTATGT TTTCTTTTCC TTCTTATGAG

1301  TTCTTTGAAA GAAAGTTCTT TGCAAGACTC GGCTCTAAGT GG
```

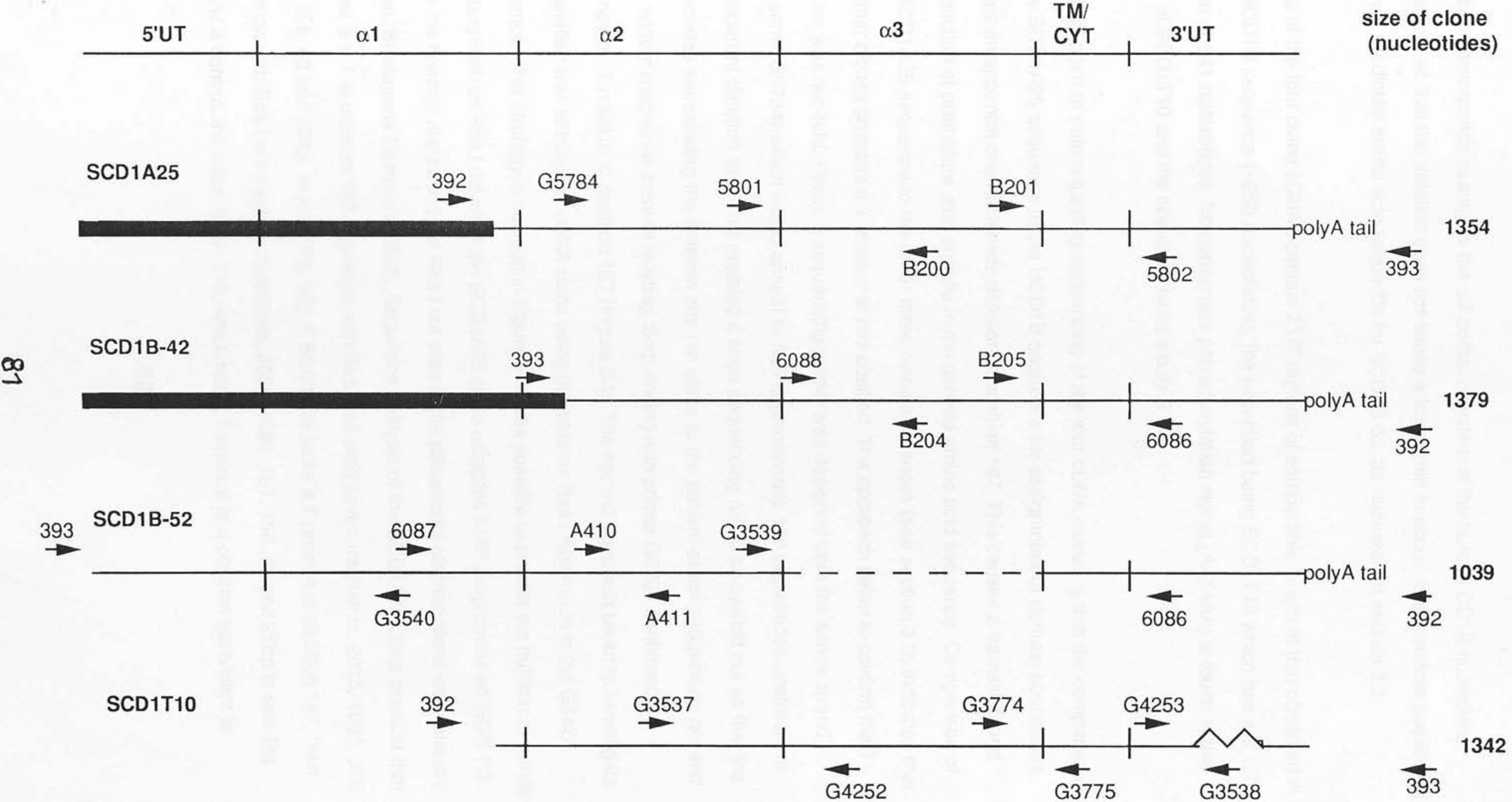
Figure 7: Sequencing Strategy and Domain Structure of The Four Ovine cDNA Clones.

The top line in the diagram is a linear representation of a CD1 molecule, with the domain structure indicated by vertical lines. The domain structure of each clone was assigned as outlined in the text. The sequencing strategy for each of the four ovine cDNA clones is shown, with the approximate annealing positions of the sequencing primers indicated with respect to the overall domain structure.

The position of each sequencing primer is shown by an arrowhead: right-facing arrows above each clone represent primer sequences derived from the sense strand, while left-facing arrows below each clone represent primer sequences derived from the antisense strand.

The sequences of primers 392 and 393 were derived from pBluescript plasmid sequence in the regions immediately adjacent to the polylinker site.

The thick black lines represent sequences which were subsequently derived by PCR. These were sequenced using M13 forward and reverse sequencing primers, as outlined in section 3.2.


5'UT
α1
α2
α3
TM/
CYT
3'UT
size of clone
(nucleotides)
SCD1A25
392
G5784
5801
B201
polyA tail
1354
B200
5802
393
SCD1B-42
393
6088
B205
polyA tail
1379
B204
6086
392
393
SCD1B-52
6087
A410
G3539
polyA tail
1039
G3540
A411
6086
392
SCD1T10
392
G3537
G3774
G4253
1342
G4252
G3775
G3538
393

deletion corresponds exactly to the $\alpha3$ domain borders of the human CD1B nucleotide sequence, so that the deletion does not cause a frameshift to occur. It is therefore possible to derive a predicted amino acid sequence for SCD1B-52, as illustrated in section 3.3.

Three of the four ovine cDNA's contain 3'UT regions of comparable length to that observed in the HCD1B sequence (~250 nucleotides), the exception being SCD1T10 which has a 3' UT region of 641 nucleotides. No consensus polyadenylation signal (AATAAA) is found in the 3'UT of SCD1T10 and the sequence lacks a poly A tail.

The final point to note regarding sequencing of the four cDNA clones is that the comparison of the SCD1A25 sequence to the HCD1B sequence for assignment of domain boundaries reveals an apparent one nucleotide deletion at position 187. This causes a frameshift and introduction of premature stop codons in the derived amino acid sequence. Comparison of the SCD1A25 sequence to the other three ovine sequences (see section 3.3) indicates that the other clones possess a T residue at this position. The approach taken to confirm the T deletion was two-fold. Firstly, a sequencing primer was designed from the sense strand sequence (G5784) which would anneal to A25 approximately 100 nucleotides upstream of the apparent deletion site. This enabled a short sequencing run to be carried out so that the nucleotides surrounding the deletion site ran close to the bottom of the sequencing gel and were widely spaced for ease of reading. Sequencing with primer G5784 confirmed the absence of T residue at position 187 (Figure 8A). The second approach taken to investigate the deletion was to map the cDNA clone using the enzyme Rsa I which cuts in the GTAC sequence. This strategy is outlined in Figure 9. It was possible to predict the number of bands resulting from an Rsa I digest of the SCD1A25 clone using the MAP programme on GCG 7.3, since the number and position of Rsa I cut sites in the pBluescript plasmid alone were already known (Stratagene Literature,1992). Sequence analysis of the SCD1A25 clone predicts that if there is a T at position 187, digestion with Rsa I will yield seven fragments, 2013, 1090, 375, 197, 156, 69 and 25bp in size (Fig. 9B). If SCD1A25 lacks a T residue at position 187, then digestion with Rsa I will yield six fragments, 2088, 1090, 197, 156, 69 and 25bp in size Fig. 9B). As a control, the clone SCD1T10, which has a T residue at a position equivalent to

Figure 8: Analysis of the SCD1A25 Clone to Determine the Presence or Absence of a T Residue at Position 187.

A. The primer G5784 was used to sequence the SCD1A25 clone, in order to verify the absence of a T residue at position 187. The sequence shown here reads from the bottom as follows, with the dot indicating the position where the T residue is missing:

CATCATTCAG.ACCATGCTATCTGTGATACT

B. The photographs show the results of Rsa I digests of various plasmid DNA samples analysed by gel electrophoresis.

Lane 1. undigested pBluescript plasmid DNA.

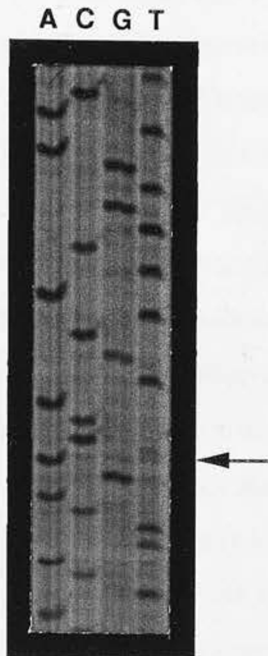
Lane 2. Rsa I digested pBluescript DNA.

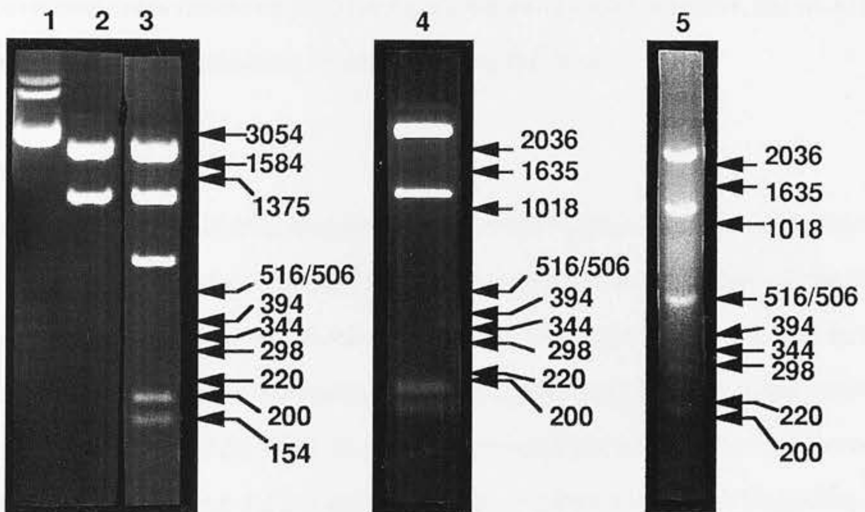
Lane 3. Rsa I digested DNA from clone A23 (identical to clone A25 by initial sequencing).

Lane 4. Rsa I digested DNA from clone A25.

Lane 5. Rsa I digested DNA from clone T10.

Molecular weight markers are indicated in base pairs.


A
A C G T


B
1 2 3
3054
1584
1375
516/506
394
344
298
220
200
154
4
2036
1635
1018
516/506
394
344
298
220
200
5
2036
1635
1018
516/506
394
344
298
220
200

nucleotide 187 in SCD1A25, was also digested with Rsa I. Sequence analysis predicts that digestion of SCD1T10 with Rsa I will yield eight fragments, 2028, 1090, 504, 285, 198, 140, 40 and 20bp in size (Fig. 9C). In all the digests, fragments of ~100bp and less were not visible in the gel. Undigested and digested pBluescript DNA samples were analysed as controls (Fig. 8B, lanes 1 and 2). The lack of a band of 375bp confirmed the absence of a T residue at position 187 in the A25 sequence (Fig. 8B, lane 4). Similarly, the A23 clone, which is identical to SCD1A25 according to the initial sequencing data, also lacked the definitive 375bp band (Fig. 8B, lane 3). However, the 2388bp fragment present in the SCD1A25 digest is absent in the digest of A23 DNA. Instead, two bands of ~1800bp and 600bp are present which appear to have been derived from the 2388bp fragment i.e. there is an internal Rsa I site somewhere within this 2388bp fragment. This fragment is composed of 1826bp of pBluescript sequence (between positions 700 and 2526) and 562bp of insert sequence (Fig. 9B). The Rsa I digestion results for A23 indicate that the extra cut site is located either within the pBluescript sequence or near the start of the insert sequence. Given that the initial sequence data for A23 was unambiguous, the extra Rsa I site appears to be located somewhere within the pBluescript segment of the 2388bp fragment. Such a site is probably the result of a rare point mutation event which occurred during the various DNA manipulation procedures. Since the A23 result supported the absence of a T residue at position 187 in the A25 sequence, no further experiments were performed to investigate the extra Rsa I digestion site in A23. The result of the SCD1T10 digestion was as expected (Fig.8B, lane5).

3.1.5 Discussion.

Studies have been carried out to investigate the ovine CD1 family at the molecular level. A human CD1C $\alpha 3$ domain probe was used to screen several sheep thymocyte cDNA libraries and the ovine HCD1B-like clone SCD1A25 was isolated from the foetal thymocyte library. A homologous $\alpha 3$/TM/CYT probe derived from this clone identified a further three ovine clones- SCD1B-42, SCD1B-52 and SCD1T10, all of which are most homologous to the human CD1B gene. Three of the four clones are truncated at the 5' end with sequences beginning around the end of the $\alpha 1$ domain or start of the $\alpha 2$ domain. The identification of 5' truncated clones is generally considered to be a feature of poor reverse transcriptase activity during library

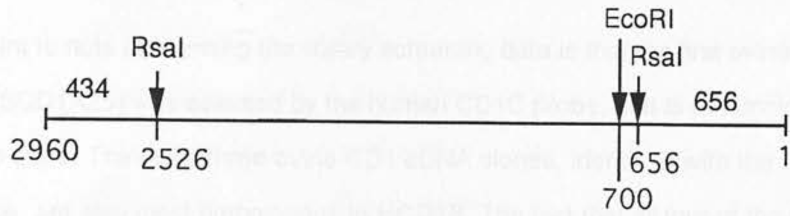
Figure 9: Rsa I Restriction Enzyme Maps of pBluescript, SCD1A25 and SCD1T10.

The three diagrams illustrate the second approach taken to verify the absence of a T nucleotide at position 187 of the SCD1A25 sequence, by mapping the clone with the restriction enzyme Rsa I, which cuts at the sequence GTAC. The diagrams show linearised versions of circular plasmids.

Diagram A illustrates the positions of the Rsa I sites in the pBluescript plasmid vector alone, as well as the location of the EcoR I site into which the lambda inserts were subcloned. Figures in plain type above the line represent distances in base pairs. Figures in alternative type below the line indicate the position of enzyme digestion sites or the termini of the plasmid sequence. Diagrams B and C are annotated in a similar manner.

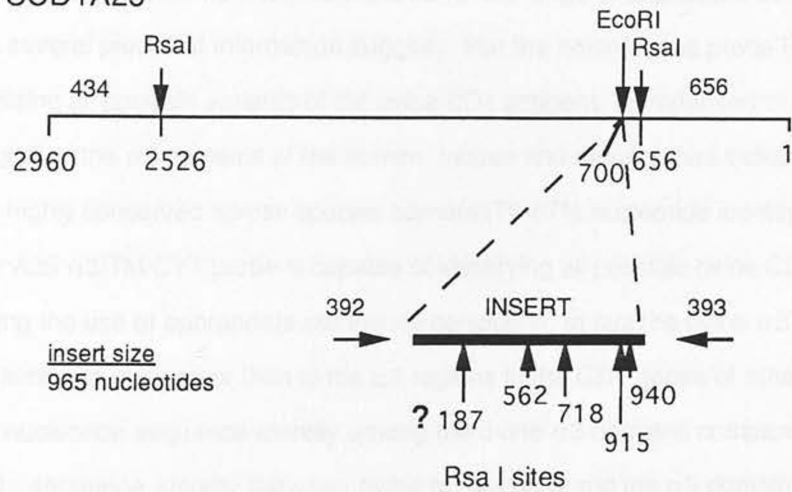
Diagrams B and **C** illustrate the orientation of the A25 and T10 inserts after subcloning, as revealed by initial sequencing with the primers 392 and 393. The positions of the Rsa I sites are marked by arrowheads as in digram A, and the predicted number and sizes of bands obtained following Rsa I digestion are shown. The SCD1T10 clone was analysed as a control,since it contained a T residue at an equivalent position to the one missing in the A25 sequence.

A pBluescript



Expected Band Sizes Following Rsa I Digestion:
1870, 1090 (434+656)

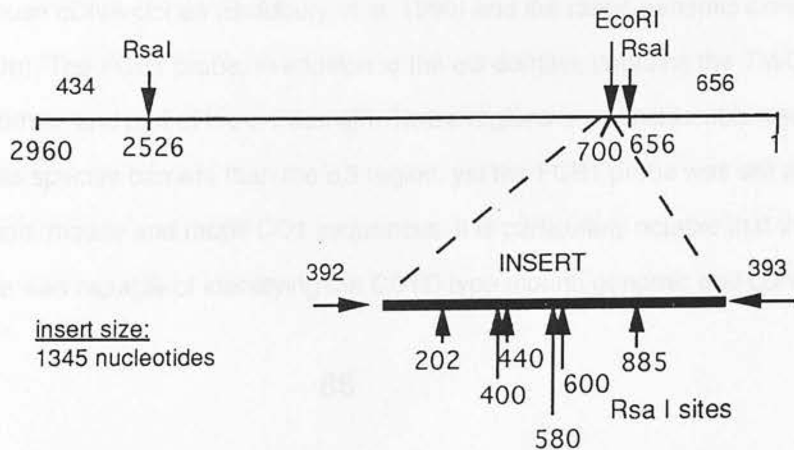
B SCD1A25



Expected Band Sizes Following Rsa I Digestion:

1. without an Rsa I site at position 187 in the insert-
   2388, 1090, 197, 156, 69, 25.
2. with an Rsa I site at position 187 in the insert-
   2013, 1090, 375, 197, 156, 69, 25.

C SCD1T10



Expected Band Sizes Following Rsa I Digestion:
2028, 1090, 504,285,198,140,40,20.

preparation. The fourth clone, SCD1B-52, represents an $\alpha3$ deletion transcript.

An interesting point to note concerning the library screening data is that the first ovine CD1 cDNA identified (SCD1A25) was selected by the human CD1C probe, and is a homologue of the human CD1B gene. The other three ovine CD1 cDNA clones, identified with the homologous probe, are also most homologous to HCD1B. The fact that all four of the ovine cDNA's are HCD1B homologues is rather unexpected given the existence of five different CD1 subtypes in humans (Martin et al, 1986). Subsequent Southern blot analysis (see section 3.4) has revealed the existence of up to seven ovine CD1 genes, and it may be that the four cDNA's identified here do not represent the complete range of expressed ovine CD1 genes. However, several pieces of information suggest that the homologous probe is capable of recognising all possible variants of the ovine CD1 antigens. Comparison of the SCD1A25 $\alpha3$ region to the $\alpha3$ domains of the human, mouse and rabbit genes indicates that the $\alpha3$ domain is highly conserved across species barriers (76-87% nucleotide identity). This suggests that the A25 $\alpha3$/TM/CYT probe is capable of identifying all possible ovine CD1 subtypes, assuming the use of appropriate stringency conditions. In fact the ovine $\alpha3$ regions are more similar to each other than to the $\alpha3$ regions in the CD1 genes of other species (90-92% nucleotide sequence identity among the ovine $\alpha3$ domains compared to 76-87% nucleotide sequence identity between ovine $\alpha3$ domains and the $\alpha3$ domains of human, rabbit and mouse genes). In a similar situation, the first human CD1 cDNA clone identified (FCB6/FCB1-two HCD1A sequences but with FCB1 extending a further 180 nucleotides into the $\alpha1$ domain; Calabi and Milstein, 1986), which contains the $\alpha3$ domain as well as part of the $\alpha2$, the TM/CYT and the 3'UT domains, was used to identify all five of the human genomic clones (Martin et al, 1986), the two mouse genomic clones (Bradbury et al, 1988), several mouse cDNA clones (Bradbury et al, 1990) and the rabbit genomic clones (Calabi et al, 1989b). The FCB1 probe, in addition to the $\alpha3$ domain, contains the TM/CYT domain, the $\alpha2$ domain and part of the $\alpha1$ domain. These regions are considerably less homologous across species barriers than the $\alpha3$ region, yet the FCB1 probe was still able to identify other human, mouse and rabbit CD1 sequences. It is particularly notable that the FCB1 CD1A probe was capable of identifying the CD1D-type murine genomic and cDNA

clones, given that the nucleotide identity between the human and murine sequences for the $\alpha1/\alpha2/\alpha3$/TM/CYT/3'UT regions, as in the FCB1 probe, is only ~55%. It seems reasonable to conclude that the A25 $\alpha3$/TM/CYT probe is potentially capable of identifying all possible variants of the ovine CD1 molecules. Likewise, the human probes are probably capable of identifying the transcribed ovine genes, since comparison of the HCD1A and HCD1C $\alpha3$ domains with the ovine $\alpha3$ domains gives an identity of 78-84%.

For the initial library screen of the SFT library, duplicate filters were taken from the eight primary screen plates. One set of filters was probed with the human CD1A $\alpha3$ probe and the second identical set screened with the human CD1C $\alpha3$ probe. In view of the fact that there is 79% nucleotide sequence identity between the $\alpha3$ domains of HCD1A and HCD1C (Calabi et al, 1991), it was expected that a significant overlap of plaques identified by each probe in the separate screenings would be observed. However, alignment of the duplicate filters revealed that few, if any, plaques were identified by both probes. This result gave the first indication of the problems which followed in screening the sheep libraries with the human probes.

A direct comparison of the results from the SFT library screen with the heterologous HCD1A and HCD1C probes and the homologous A25 $\alpha3$/TM/CYT probe provides a concise overview of the problems encountered during the screening work. When the SFT library was screened with the HCD1C probe (series A), thirty to fifty plaques were identified per 30,000 plaques screened. A similar frequency of positive clones was observed when the same library was screened with the HCD1A probe (series B). However, of the six EcoR I excisable inserts obtained from twenty-five positive plaques isolated from the secondary screen of the A series (1.0-1.2kb in size), only two were CD1 equivalents, and these were identical. SCD1A25 was retained for further analysis. Following Mlu I digestions of the positive clones which did not yield Eco RI excisible inserts, several inserts were subcloned and four were sequenced, although none were found to be homologous to CD1. For the series B clones, seventeen EcoR I excisable inserts were obtained from a total of twenty-eight positive plaques isolated from the secondary screen. In contrast to the series A clones, several of the series B clones contained double inserts, and most bands were in the size range 3-4kb. The size of these

inserts gave some cause for concern, since 3kb is the maximum size of any CD1 transcript observed so far (Calabi et al, 1988). None of the inserts turned out to be CD1 homologues. Screening of the SFT library with the A25 α3/TM/CYT probe, although not a direct contrast to the results obtained by screening with the heterologous human probes because of the differences in stringency conditions, gave a frequency of twenty positive plaques per 150,000 plaques screened. One EcoR I excisable insert (T10, ~1.3kb) was obtained from twelve positive plaques identified in the secondary screen, and initial sequence analysis confirmed that it was a CD1 homologue.

There are two main points to note regarding the screening results outlined above. Clearly, the heterologous probes were much less specific than the homologous probe at identifying ovine CD1 clones, as indicated by the numbers of positive plaques identified. This was confirmed by screening the ST-1 and ST-2 libraries with the homologous probe. Although not a direct comparison with the SFT library results, these screenings gave a similar frequency of positive plaques. Given the high percentage nucleotide identity between the α3 domains of the human, mouse, rabbit and ovine sequences (76-86%), the results seem to suggest that the flanking regions contained within the M13 human clones (~0.3kb of flanking intron sequence from a total insert size of 0.6kb) may have significantly contributed to the screening difficulties encountered. With hindsight it appears that the initial library screens might have been more successful had most of the flanking region sequence been removed. For example, the PCR technique could have been used to amplify the α3 region from the M13 DNA using primers based on the terminal sequences of the α3 domain.

The second point to note is that library screening with the two human probes gave very different results. The human CD1C probe was derived by PCR amplification from the original M13 clone, and was confirmed as CD1C by sequencing analysis. The human CD1A probe was obtained by directly subcloning the original M13 insert without PCR amplification. Since this probe was not confirmed as CD1A by sequence analysis, it is possible, though unlikely, that something other than a CD1 sequence was subcloned and subsequently used for library screening. An alternative explanation for the differences observed in screening with the two

human probes would be that the flanking intron sequences for the two probes differentially contributed to the identification of false positives.

A third minor point to note regarding the screening results is that, with hindsight, the PCR technique could have been used to amplify inserts identified as positive in secondary screens but which did not yield Eco RI excisable inserts. This would have enabled subcloning and sequencing of all positive samples to be completed.

Sequence analysis of the SCD1A25 clone revealed an apparent one nucleotide deletion at position 187 in the sequence, introducing premature stop codons in the deduced amino acid sequence (see section 3.3). This was initially considered to be due to error in reading the sequence, but further experiments ruled out this possibility. The deletion was verified by sequencing with the primer G5784 and by Rsa I digestion of the cloned DNA, as described previously. It is possible that the deletion represents a cloning artefact which could have arisen during synthesis of the cDNA used to prepare the library. This possibility could be investigated by isolation and sequencing of a genomic clone equivalent to SCD1A25, or by *in vitro* translation of A25 mRNA isolated using the cloned cDNA. In the latter case, the translation product is analysed by gel electrophoresis to determine its size and thus confirm the presence or absence of premature stop codons in the nucleotide sequence.

The two clones SCD1B-42 and SCD1B-52 were both isolated from the ST-2 library. SCD1B-42 contains a truncated CD1, while the SCD1B-52 insert lacks the region which corresponds exactly to the $\alpha3$ domain. Comparisons of these two sequences at the nucleotide level showed that they are ~96% identical. A more detailed discussion of the implications of this high percentage sequence identity between SCD1B-42 and SCD1B-52 is presented in section 3.3.5. The SCD1B-52 $\alpha3$ deletion clone was presumably identified by the A25 $\alpha3$/TM/CYT probe as a result of interaction between the TM/CYT domains of the probe and the SCD1B-52 sequence. Comparison of the TM/CYT region of SCD1B-52 and SCD1A25 gives 88.5% identity over 113 nucleotides. No comparable deletion to that observed in SCD1B-52 has been detected among any of the other known CD1 mRNA splice variants in

humans and mice (Calabi et al, 1989c; Bradbury et al, 1990; Woolfson and Milstein, 1994), although Kirszenbaum et al (1994) have recently reported an alternatively spliced form of the human nonclassical class I molecule HLA-G which lacks exon 4 (corresponding to the $\alpha 3$ domain).

Both cDNA sequence analysis and nuclease protection assays suggest the existence of a complex array of alternatively spliced human CD1 transcripts (Calabi et al, 1989c). Cryptic splice sites have been detected in the leader-$\alpha 1$ intron of HCD1A, in the leader exon and the leader-$\alpha 1$ intron of HCD1E, and within the $\alpha 3$ exons of HCD1B and HCD1E (Calabi et al, 1991). Most of these splicings cause frameshifts leading to chain termination or probable destruction of β_2m binding and membrane expression, whereas the SCD1B-52 sequence appears to be translatable. A recent analysis of human CD1 expression in mouse myeloma transfectants and in human thymocytes has revealed further mRNA splicing complexity (Woolfson and Milstein, 1994). Mouse myeloma cells were transfected with a partially spliced CD1A cDNA and the 10B3 clone selected for further analysis. Total 10B3 mRNA was copied into cDNA and used as a template in PCR reactions with CD1A primers from the $\alpha 3$ and 3' untranslated regions. Four bands were detected, cloned and sequenced. One PCR band was found to correspond to an unspliced transcript with a cryptic stop codon in the intron between the $\alpha 3$ and TM/CYT domains. Protein sequencing of CD1a purified from the transfectant supernatant indicated that the unspliced transcript was a secretory isoform of CD1a. A second PCR band was found to correspond to an alternatively spliced transcript in which a cryptic splice acceptor site within the transmembrane/cytoplasmic exon is utilised in preference to the site at the beginning of the exon. The product of this transcript was found to be an intracellular component. A third band represented the correctly spliced membrane-bound form of CD1a and the fourth band was a heteroduplex derived from two of the other bands. A similar analysis of all five HCD1 genes was performed in thymus tissue. A single major band was detected following amplification of CD1A and CD1B cDNA, of a size consistent with the correctly spliced membrane-bound form of CD1. The single CD1A transcript detected in thymus tissue contrasts sharply with the complex splicing pattern observed in the CD1A transfectant which was possibly due to overproduction. CD1D amplification gave a major band

but in addition gave two minor bands. In contrast, CD1C and CD1E gave complex splicing patterns. The pattern for CD1C was comparable to that observed in the CD1A transfectant, with unspliced, alternatively spliced and correctly spliced transcripts being detected. Human CD1 gene expression is thus characterised by considerable mRNA splicing complexity which may be tissue specific.

In the mouse, analysis of cDNA clones from total thymus RNA has revealed that the majority derive from partially spliced transcripts (Bradbury et al, 1990; Bradbury, A. PhD Thesis). The problems and implications of alternative splicing of mouse CD1 transcripts from thymus tissue have been fully discussed by Bradbury et al (1990). These authors highlight the fact that alternative splicing is not unique to the CD1 genes in the thymus and neither is it unique for thymus tissue. Introns of different genes have been found to be removed with different efficiency in different tissues. In fact, RNase mapping experiments with probes containing each of the domains of mouse CD1D with a section of adjacent intron, showed that for mouse CD1 there was a significant degree of unspliced message (20-50% of total) for the $\alpha1$, $\alpha2$ and $\alpha3$ regions in the thymus, a small degree in the spleen (1-5%) and normal splicing in the liver (Bradbury, A. PhD Thesis). The significance of the predominance of alternatively spliced CD1 isoforms in the thymus is unclear, although it is of interest that thymic CD1 expression is confined to immature cortical thymocytes and is negatively correlated with MHC class I expression (Calabi et al, 1991). Heterogeneity of mRNA transcripts has been detected by Northern blot analysis for the mouse class Ib gene TLa, with transcript sizes ranging from 1.3-3.5 kb (Flavell et al, 1986). Alternative splicing events have also been detected for the Qa-2 class Ib gene (Tabaczewski et al,1994), for ovine and other classical class I molecules (Grossberger et al, 1990; Guillemot et al, 1988), and for a wide variety of other mammalian genes (Breitbart et al, 1987). Alternative splicing of CD1 mRNA may provide a mechanism for regulating gene expression (Calabi et al, 1991).

Many of the known alternative splicing events for the CD1 molecules represent the use of alternative donor and acceptor sites for splicing which lie within potential coding sequence, and result in mRNA species which cannot be translated. In contrast to this, the $\alpha3$ deletion of

SCD1B-52 represents exclusion of a discrete exon of genetic information, which would result in the inability of the translated product to interact with β2m. The biological significance of such a molecule is unclear, and it may be functionally unimportant. However, when alternatively spliced message has been identified from MHC class I genes, the cognate protein has often been subsequently found (Rodgers et al, 1986; McCluskey et al, 1986; Lew et al, 1987). The potential for membrane expression of an SCD1B-52 protein product remains to be established, since it is unclear whether interaction with β2m, which is necessary for the correct folding and membrane expression of MHC class I α chains (Zijlstra et al, 1990), is an absolute requirement for surface expression of all CD1 molecules (Calabi et al, 1991). Although the α chains of HCD1a and HCD1b can be expressed on the cell surface independently of β2m, this appears to be the result of dissociation from β2m at the cell surface (Calabi et al, 1991). Interestingly, Balk et al (1994) have recently reported the expression of HCD1d on human intestinal epithelial cells as a 37kD, non-β2m associated protein with no N-linked carbohydrate. Transfection into a β2m-negative cell line confirmed that HCD1d (and the murine equivalent, CD1.1) can be expressed at the cell surface in the absence of β2m. It is not known whether such β2m-independent CD1d molecules are functionally significant or not. The authors suggest that the lack of association with β2m may be dictated by the conformation of the non-glycosylated CD1d α chain, indicating that a conventional site for peptide binding may not be present. Similarly, the protein encoded by SCD1B-52 may be conformationally different compared to typical CD1 molecules, and may therefore not possess a conventional class I-like peptide binding groove. The question of structural differences between an SCD1B-52 translation product and conventional CD1 molecules could be resolved by elucidation of their crystal structures and comparison to the known crystal structures of MHC class I molecules.

A clone (A-14) identified by initial sequencing to be identical to SCD1B-52, was also isolated from the adult ST-1 library. However, it was unlikely that A-14 contained the α3 region either, since the original lambda insert was identical in size and by initial sequencing data to that observed for the SCD1B-52 clone. The identification of an α3 deletion cDNA clone from two different libraries supports the view that the existence of such sequences is not due to a cloning artefact.

Several potential experiments could be performed in order to further define the nature of SCD1B-52 transcription and expression. The transcription of SCD1B-52 could be investigated using a sensitive hot-start PCR technique, as employed in the investigation of HLA-G transcription (Kirszenbaum et al, 1994), using primers specific for the $\alpha2$ and 3'UT regions of SCD1B-52. Alternatively, Northern hybridisation using a probe based on the sequence of the $\alpha2$/TM border would enable the transcription of SCD1B-52 to be investigated in a variety of tissues. This probe could also be used to confirm the specificity of any products amplified by PCR. It may also be possible to study SCD1B-52 membrane expression by generating an antipeptide antiserum against an epitope encoded by the junction of the $\alpha2$ and TM nucleotide sequences. This latter method has recently been successfully employed to study the alternatively spliced transcripts of the nonclassical murine class I molecule Qa-2 (Tabaczewski et al, 1994). An antipeptide antiserum was generated against an epitope encoded by the junction of exon 4 and exon 6 in Qa-2 which was able to distinguish soluble Qa-2 molecules produced from exon 5 (TM domain) deletion transcripts and those derived from membrane-bound antigens.

The SCD1T10 clone contained a truncated CD1 with a long 3'UT region of 641 nucleotides as compared to 260-280 nucleotides for the 3'UT's of the other clones. The insert lacked a poly A tail and no polyadenylation signal (AATAAA) was observed. This is a similar situation to that observed after the cloning of HCD1D cDNA, which has a 3'UT of 705 base pairs with no poly A tail or polyadenylation signal (Balk et al, 1989). In the latter case, sequencing of the corresponding genomic clone revealed a probable polyadenylation signal 21 base pairs downstream from the end of the cDNA sequence. It was noted that the cDNA library from which SCD1T10 was isolated (SFT), had been generated by isolation of messenger RNA on a poly dT column, followed by cDNA synthesis (Hein et al, 1989). It was therefore assumed that the SCD1T10 sequence, by virtue of being present in the cDNA library, had possessed a poly A tail at least until the point of being selected on the poly dT column. Messenger RNA transcripts lacking poly A tails are very unstable (Wickens et al,1990; Wahle and Keller,1992), and it is unlikely that such a transcript would have survived selection on the poly dT column and cDNA synthesis. Careful examination of the region by alignment of all the ovine

sequences revealed a potential polyadenylation signal in an equivalent site to those observed in the other clones, although in this case the sequence read ACTAAA. The 3' UT for SCD1T10 extended a further 379 nucleotides downstream of this site, whereas the poly A tail in the other clones is located just 15 nucleotides downstream of the polyadenylation signal. Comparison of the nucleotide sequences preceding the polyadenylation sites in a wide variety of other eukaryotic mRNA's has revealed the presence of the ubiquitous hexanucleotide sequence AAUAAA (Wahle and Keller,1992). Mutagenesis experiments have since confirmed that this consensus sequence, typically located 10-35 nucleotides upstream of the polyadenylation site, is essential for 3'-end formation of polyadenylated mRNA. The only common natural variant is AUUAAA. Comparison of the efficiency of polyadenylation of a variety of mutant polyadenylation sites has shown that the sequence AAUAAA is 100% efficient, and the most common variant, AUUAAA is 80% efficient (Wickens et al, 1990). The sequence found in SCD1T10 (ACUAAA for the RNA equivalent) was shown to be only 10% efficient for polyadenylation of the mRNA. It is apparent that the ACTAAA sequence in SCD1T10 was not used as a polyadenylation signal particularly since isolation by library screening from a cDNA library indicates that the transcript must be quite abundant in thymus tissue *in vivo*. This does not however rule out the possibility that some naturally occurring T10 transcripts may be able to use the ACTAAA sequence for induction of polyadenylation. It seems that the polyadenylation signal and poly A tail of SCD1T10 have been lost at somepoint during library generation. Isolation and sequencing of the genomic clone equivalent to SCD1T10 cDNA, as for HCD1D, would confirm the use of an alternative polyadenylation signal downstream to that used by the other ovine sequences.

3.2 PCR Amplification of 5'UT/α1 Regions.

3.2.1 Introduction

Of the four ovine CD1 cDNA clones isolated by library screening, three were found to contain inserts truncated at the 5' end. The SCD1A25, SCD1T10 and SCD1B-42 inserts lack the 5'UT and α1 domains. The isolation of 5' truncated sequences from cDNA libraries is a relatively common phenomenon and is considered to reflect poor reverse transcriptase activity during cDNA synthesis for library preparation. The PCR technique was employed to isolate part of the 5' untranslated region as well as the α1 domain of the truncated clones in order to obtain full length sequence data encoding the ovine CD1 molecules. The region amplified is referred to as the "5'UT/α1 region" throughout this section.

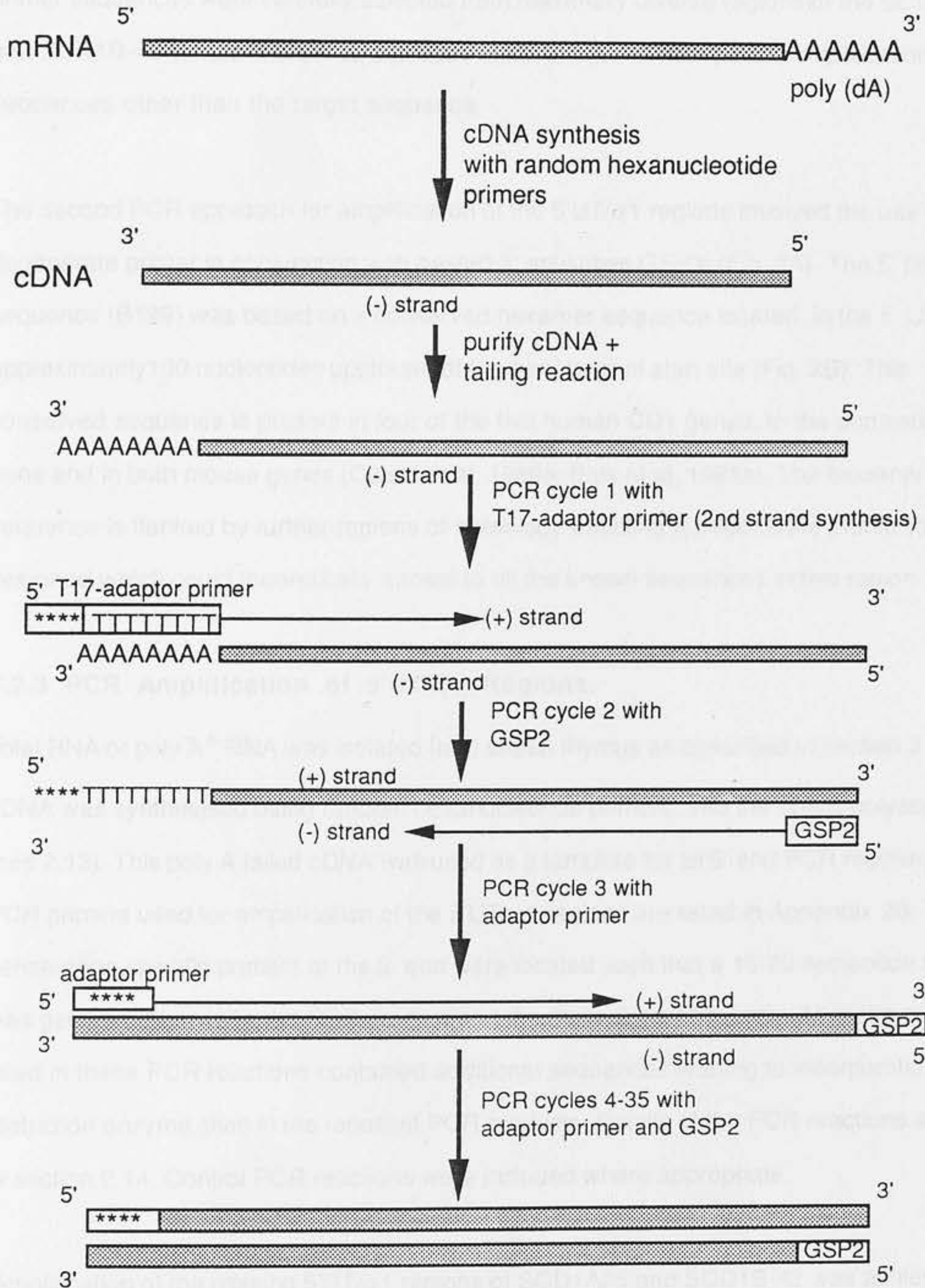
3.2.2 Experimental Rationale for PCR Of The 5'UT/α1 Regions.

Two main PCR approaches were employed to amplify the unknown 5'UT/α1 regions of the ovine CD1 sequences. The first approach involved a modification of the anchored PCR or RACE (Rapid Amplification of cDNA Ends) PCR technique (Loh et al, 1989; Fig.1). In essence, the 5' RACE protocol generates cDNA molecules by performing reverse transcription using a gene specific primer (GSP). The specific first strand cDNA is polyadenylated using terminal deoxy nucleotidyl transferase (TdT) and dATP. The poly A tract serves as the target sequence for a hybrid poly T containing primer in the subsequent amplification process. The hybrid primer consists of oligo (dT) linked to a unique 17-base oligonucleotide "adaptor" primer. One advantage of using a poly A tail concerns the fact that A-T binding is weaker than G-C binding, so that longer stretches of A residues in the cDNA are required for efficient nonspecific binding of the hybrid primer (Rolfs et al, 1992). The unknown 5'UT/α1 region is amplified using the hybrid primer and a second GSP upstream of the first one. One drawback of this approach is that specifically primed first strand cDNA must be synthesised for each transcript under study. A second problem with RACE is that specific primers may sometimes anneal inefficiently to the transcript of interest as a result of mRNA secondary structure. Random primed cDNA synthesis in conjunction with RACE can be employed to overcome such difficulties (Harvey and Darlinson, 1991) and was used in this

Figure 1: RACE-PCR of 5'UT/α1 Regions.

Schematic representation of the RACE protocol for PCR amplification of 5'UT/α1 regions. The diagram is simplified at each stage to illlustrate only how the new product formed during the previous step is utilised.

The diagram illustrates cDNA synthesis and the first round of PCR utilising the hybrid T17-adaptor primer (B186), the adaptor primer (B185) and the external gene specific primer (GSP2). The subsequent nested PCR reaction (not shown) utilised the adaptor primer and the internal gene specific primer (GSP1) only. All primer sequences are listed in Appendix 2B, and the PCR protocol is detailed in section 2.14.



5'
mRNA
3'
AAAAAA
poly (dA)
cDNA synthesis
with random hexanucleotide
primers
3'
cDNA
5'
(-) strand
purify cDNA +
tailing reaction
3'
AAAAAAAA
5'
(-) strand
PCR cycle 1 with
T17-adaptor primer (2nd strand synthesis)
5' T17-adaptor primer
****TTTTTTTT
(+) strand
3'
AAAAAAAA
3'
(-) strand
5'
PCR cycle 2 with
GSP2
5'
****TTTTTTTTT
(+) strand
3'
3'
(-) strand
GSP2
5'
PCR cycle 3 with
adaptor primer
adaptor primer
5'
****
(+) strand
3'
3'
GSP2
(-) strand
5'
PCR cycles 4-35 with
adaptor primer and GSP2
5'
****
3'
GSP2
3'
5'

case. The use of nested antisense 3' GSP's is known to greatly increase the specificity of the RACE technique (Rolfs et al, 1992) and was also employed in this case. The antisense 3' primer sequences were carefully selected from maximally diverse regions of the SCD1A25 and SCD1B-42 clones in order to decrease the likelihood of nonspecific amplification of CD1 sequences other than the target sequence.

The second PCR approach for amplification of the 5'UT/α1 regions involved the use of a 5' degenerate primer in conjunction with nested 3' antisense GSP's (Fig. 2A). The 5' primer sequence (B199) was based on a conserved hexamer sequence located in the 5' UT region approximately100 nucleotides upstream of the translational start site (Fig. 2B). This conserved sequence is present in four of the five human CD1 genes, in the domestic rabbit gene and in both mouse genes (Calabi et al, 1989a; Balk et al, 1991a). The hexamer sequence is flanked by further regions of homology enabling a degenerate primer to be designed which could theoretically anneal to all the known sequences in this region.

3.2.3 PCR Amplification of 5'UT/α1 Regions.

Total RNA or poly A^+ RNA was isolated from sheep thymus as described in section 2.12. cDNA was synthesised using random hexanucleotide primers, and the cDNA polyadenylated (see 2.13). This poly A tailed cDNA was used as a template for all 5' end PCR reactions. The PCR primers used for amplification of the 5'UT/α1 regions are listed in Appendix 2B. The anti-sense gene specific primers at the 3' end were located such that a 15-20 nucleotide overlap was generated between the PCR product and the main clone sequence. All of the primers used in these PCR reactions contained additional sequences leading to incorporation of restriction enzyme sites in the resultant PCR products. Details of the PCR reactions are given in section 2.14. Control PCR reactions were included where appropriate.

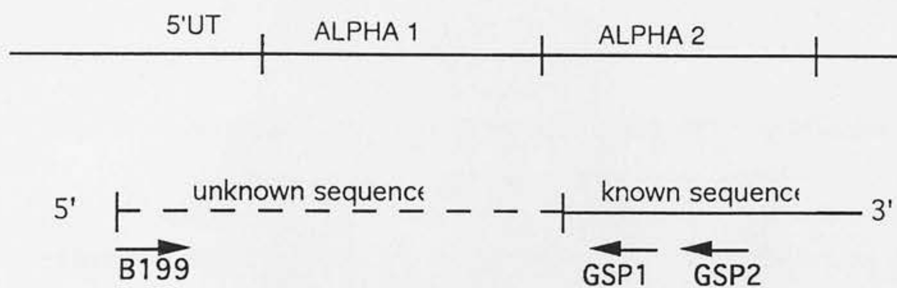
Amplification of the missing 5'UT/α1 regions of SCD1A25 and SCD1B-42 was achieved using both PCR strategies described above (Fig. 3). The PCR products generated were of the expected size (~450-500 nucleotides). The attempted amplification of the 5'UT/α1 region of the SCD1T10 sequence was unsuccessful.

Figure 2: PCR of The 5'UT/α1 Regions Utilising The Degenerate 5' Primer B199.

A. The top line in thediagramis a linear representation of a CD1 molecule with the domain structure indicated by vertical lines. The lower part of the diagram shows a basic outline of the 5'UT/α1 PCR technique utilising the degenerate 5' primer B199 and the nested 3' gene specific primers.

B. The aligned sequences illustrate the derivation of the 5' degenerate primer B199. The conserved hexamer sequence GAAGTC is located ~100 nucleotides upstream of the translation initiation codon in the 5' UT region, and is flanked by semiconserved regions of sequence. Outwith the hexamer sequence, nucleotides were preferentially selected from the HCD1A, B, C and DomRab sequences, since the HCD1E, MCD1.1 and MCD1.2 sequences belong to a different class of CD1 molecules (Calabi et al, 1991).

A



1st round of PCR: B199 + GSP2 (external gene specific primer)
2nd round of PCR: B199 + GSP1 (internal gene specific primer)

B

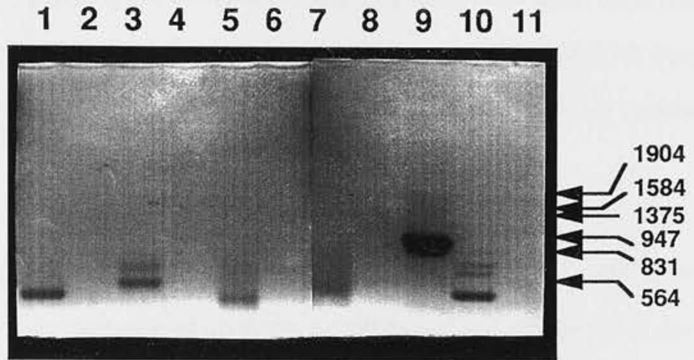
```
                -98                  -79
  HCD1A      GAAAGAAGTCAGAATAGAG
  HCD1B      AGAAGAAGTCACTACAGGG
  HCD1C      GAAGGAAGTCAGAATATAG
  HCD1E       AGGGAAGTCAGACGA
 DomRab      AGAGGAAGTCACTACAGAG
 MCD1.1      CGCAGAAGTCGGAGCCGAG
 MCD1.2      TGCAGAAGTCGGAACCCAG


          5'                     3'
             AA A       GA C
   B199 ***    G GAAGTCA  A AGAG
             GG G       CT T
```

Figure 3: Results of PCR Amplification of 5' UT/α1 Regions.

The photograph shows the results of PCR amplification of the 5'UT/α1 regions of SCD1A25 and SCD1B-42. 20% of the PCR products from the nested reaction were analysed by gel electrophoresis in a 1.5% agarose gel.

Lane 1: 5'UT/α1 of SCD1A25 amplified using the 5' degenerate primer B199 in conjunction with the nested 3' GSP's.
Lane 3: 5'UT/α1 of SCD1B-42 amplified using the 5' degenerate primer B199 in conjunction with the nested 3' GSP's.
Lane 5: 5'UT/α1 of SCD1A25 amplified using the RACE-PCR protocol.
Lane 7: 5'UT/ α1 of SCD1B-42 amplified using the RACE-PCR protocol.
Lane 9: positive control PCR (amplification of the first 936 nucleotides of the env gene of Maedi Visna virus strain EV1, using primers 240X and 767F and ~5ng of plasmid pA46 DNA: Carey and Dalziel,1994) to show that all PCR reaction components were working effectively.
Lane 10: as for lane 1, but using an alternative first round PCR sample as template.
Lanes 2, 4, 6, 8 and 11: negative control PCR reactions where the DNA template was substituted by an equal volume of SDW.

Molecular weight markers are indicated in base pairs.


1 2 3 4 5 6 7 8 9 10 11
1904
1584
1375
947
831
564

Initial attempts at cloning the purified PCR products by restriction enzyme digestion and ligation into appropriately cut pBluescript were not successful for unknown reasons. The remainder of the PCR products were finally cloned using the TA cloning kit (Invitrogen). However, the small number of clones derived from the RACE PCR products were found to be false positives, probably reflecting the small amount of amplified DNA which was then available for cloning. Cloning of the PCR products resulting from the reactions using the degenerate 5' primer was much more successful, yielding several insert-positive clones for sequencing.

3.2.4 Presentation of 5'UT/α1 Sequence Data.

The inserts from five SCD1A25 5'UT/α1 region clones and three SCD1B-42 5'UT/α1 region clones were essentially fully sequenced on each strand by dsDNA sequencing using the dideoxy chain termination technique (Sanger et al,1977) and M13 forward and reverse sequencing primers (see Appendix 2A).The sequences obtained for the sense and antisense strands of each clone were compared and any ambiguities resolved by referring back to the original sequencing autoradiographs. The multiple sequences for the 5'UT/α1 regions of SCD1A25 (Fig.4A) and SCD1B-42 (Fig.4B) were aligned using the PILEUP programme of the UWGCG version 7.3 sequence analysis software package (Devereux et al, 1984). Pairwise comparisons between the SCD1A25 and SCD1B-42 sequences were performed using the GAP programme of UWGCG 7.3.

The 1A25, 3A25 and 5A25 5'UT/α1 region sequences are identical outwith the degenerate 5' primer site except for two single base differences over 1347 (449x3) nucleotides (Fig. 4A). The 3A25 sequence contains an A residue at position 331 compared to a G residue present in the other two sequences. The 5A25 sequence contains a G residue at position 378 as opposed to an A residue at the same position in the other two sequences. There are sixteen nucleotide differences outwith the degenerate 5' primer site between the final two clones, 2A25 and 4A25, as well as a four nucleotide deletion in the 5' UT region of 4A25. The single nucleotide differences are located at positions 33, 34, 76, 79, 101, 102, 110, 111, 165, 210, 273, 280, 287, 370, 383 and 387.

Figure 4: Sequences of The 5'UT/α1Regions of SCD1A25 and SCD1B-42.

The sequences illustrated in Figure 4 represent the sequence information derived by dsDNA sequencing of each strand of the five SCD1A25 5'UT/α1 clones (Fig. 4A) and the three SCD1B-42 5'UT/α1 clones (Fig.4B).

Domain boundaries are indicated by vertical dashed lines.

Long sequences (~20 nucleotides) marked in italic bold type represent the primers used in the nested PCR reaction.

The trimer sequence marked in bold type (ATG) is the translation initiation codon.

Differences between the sequences are indicated by a single underline beneath the sequence alignment.

The overlap of the PCR sequences with the main clone sequence begins at nucleotide 415 for the SCD1A25 5' ends and at nucleotide 483 for the SCD1B-42 5' ends.

The horizontal line dividing the first three virtually identical SCD1A25 5' sequences from the other two illustrates the clear division of the first three sequences from the other two similar but different sequences (see text for explanation).

Fig. 4A

```
     1    B199               5'UT Domain                 50
1A25 GGAGGAAGTC ACTATAGAGT ACTGAGAAAA A..GGTTTGC TGAAATTAGA
3A25 AAAAGAAGTC ACTATAGAGT ACTGAGAAAA A..GGTTTGC TGAAATTAGA
5A25 GAAAGAAGTC AGTACAGAGT ACTGAGAAAA A..GGTTTGC TGAAATTAGA
2A25 GGAGGAAGTC AGTATAGAGT ACTGAAAAAA AAAAAAAAGC TGAAATTAGA
4A25 AGAGGAAGTC AGTACAGAGT ACTGAAAAAA AATGA....C TGAAATTAGA

     51                                                 100
1A25 GATCAAATAC CAATTCTGAG AGTCAGAAGC TCTACTTCCC ATTGAGATGC
3A25 GATCAAATAC CAATTCTGAG AGTCAGAAGC TCTACTTCCC ATTGAGATGC
5A25 GATCAAATAC CAATTCTGAG AGTCAGAAGC TCTACTTCCC ATTGAGATGC
2A25 GATCAAATAC CAGCTCCGAG AGTCAGAAAT TCTACTTCCC AGTGAAATGC
4A25 GATCAAATAC CAGCTCCGAG AGTCACAAGT TCTACTTCCC AGTGAAATGC

     101    Leader Domain                              |150
1A25 TGCTTCTGCC ACTTCTATTG CTAGCAGTTA TTGTGCCAGG TGGTGAC|AAT
3A25 TGCTTCTGCC ACTTCTATTG CTAGCAGTTA TTGTGCCAGG TGGTGAC|AAT
5A25 TGCTTCTGCC ACTTCTATTG CTAGCAGTTA TTGTGCCAGG TGGTGAC|AAT
2A25 TGCTTCTACC GCTTCTGTTA CTTGGAGTTA TCCTCCCAGG TGGTGAC|AAT
4A25 CTCTTCTACT ACTTCTGTTA CTTGGAGTTA TCCTCCCAGG TGGTGAC|AAT

     151       Alpha 1 Domain                           200
1A25 GAGGATGTAT TCCAGGGGCC AACCTCCTTC CATGTCATCC AGATTTCGAC
3A25 GAGGATGTAT TCCAGGGGCC AACCTCCTTC CATGTCATCC AGATTTCGAC
5A25 GAGGATGTAT TCCAGGGGCC AACCTCCTTC CATGTCATCC AGATTTCGAC
2A25 GAGGATGTGT TCCAGGGGCC AACCTCCTTC CATCTCAAGC AGATTTCAAC
4A25 GAGGATGTGT TCCAAGGGCC AACCTCCTTC CATCTCAAGC AGATTTCAAC

1A25 ATTTGCCAAC AGCACTTGGG CTCAAAATCA AGGCTCAGGC TGGTTGGACA
3A25 ATTTGCCAAC AGCACTTGGG CTCAAAATCA AGGCTCAGGC TGGTTGGACA
5A25 ATTTGCCAAC AGCACTTGGG CTCAAAATCA AGGCTCAGGC TGGTTGGACA
2A25 CTTTGTCAAC AGCACATGGG CTCAAAATCT AGGCTCAGGC TGGTTGGATG
4A25 CTTTGTCAAT AGCACATGGG CTCAAAATCT AGGCTCAGGC TGGTTGGATG

     251                                                300
1A25 ATTTGCAGCT TTATGGCTGG GACAGTGACC CAGGCACTAC CATCTTCCTG
3A25 ATTTGCAGCT TTATGGCTGG GACAGTGACC CAGGCACTAC CATCTTCCTG
5A25 ATTTGCAGCT TTATGGCTGG GACAGTGACC CAGGCACTAC CATCTTCCTG
2A25 ACTTGCAGAT TCATGGCTGG GAGAGTGACT CGGGCACTGC CATTTTCCTG
4A25 ACTTGCAGAT TCATGGCTGG GACAGTGACC CTGGCACTGC CATTTTCCTG

     301                                                350
1A25 AAGCCCTGGT CTAAGGGCAA CTTCAGTGAT GAGGAGGTGA CTGAGCTGGA
3A25 AAGCCCTGGT CTAAGGGCAA CTTCAGTGAT AAGGAGGTGA CTGAGCTGGA
5A25 AAGCCCTGGT CTAAGGGCAA CTTCAGTGAT GAGGAGGTGA CTGAGCTGGA
2A25 AAGCCCTGGT CCAAGGGCAA CTTTAGTGAT GAGGAGATAA CTGAGCTGGT
4A25 AAGCCCTGGT CCAAGGGCAA CTTTAGTGAT GAGGAGATGA CTGAGCTGGT

     351                                                400
1A25 GGAACTATTT AGAGTCTACC TCATTGGATT TACTCTAGAA GTGCAGGACC
3A25 GGAACTATTT AGAGTCTACC TCATTGGATT TACTCTAGAA GTGCAGGACC
5A25 GGAACTATTT AGAGTCTACC TCATTGGGTT TACTCTAGAA GTGCAGGACC
2A25 GGACCTCTTC CGAGTCTACC TCATTGGATT CATTCGGGAA GTGCAGGATC
4A25 GGACCTCTTC CGAGTCTACT TCATTGGATT CACTCGAGAA GTGCAGGATC

                                  Alpha 2
     401                         | Domain    B202       449
1A25 ATGTCAGTGA ATTCCAGCTG GAAT|ACCCCT TTGTGATCCA GGACATAGC
3A25 ATGTCAGTGA ATTCCAGCTG GAAT|ACCCCT TTGTGATCCA GGACATAGC
5A25 ATGTCAGTGA ATTCCAGCTG GAAT|ACCCCT TTGTGATCCA GGACATAGC
2A25 GAGTCAATGA GTTCCAGTTA GAAT|ACCCCT TTGTGATCCA GGACATAGC
4A25 GAGTCAATGA GTTCCAGTTA GAAT|ACCCCT TTGTGATCCA GGACATAGC
```

Fig. 4B

```
      1     B199                                               50
1B-42 AAAGGAAGTC AGTACAGAGT ACTG...AAA AAAAAAAAAG CTGAAATTAG
3B-42 GAAGGAAGTC AGTACAGAGT ACTGAAAAAA AAAAAAAAAG CTGAAATTAG
2B-42 GGAAGAAGTC ACAACAGAGT ACTG....AA AAAAAATGAG CTGAAATTAG

      51         5'UT Domain                                  100
1B-42 AGATCAAATA CCAGCTCCGA GAGTCAGAAA TTCTACTTTC CAGTGAAATG
3B-42 AGATCAAATA CCAGCTCCGA GAGTCAGAAA TTCTACTTTC CAGTGAAATG
2B-42 AGATCAAATA CCAGCTCCGA GAGTCAGAAG TTCTACTTCC CAGTGAAATG

      101                                                    |150
1B-42 CTGCTTCTAC CGCTTCTGTT ACTTGGAGTT ATCCTCCCAG GTGGTGAC|AA
3B-42 CTGCTTCTAC CGCTTCTGTT ACTTGGAGTT ATCCTCCCAG GTGGTGAC|AA
2B-42 CTGCTTCTAC CACTTCTGTT ACTTGGAGTT ATCCTCCCAG GTGGTGAC|AA

      151        Alpha 1 Domain                               200
1B-42 TGAGGATGTG TTCCAGGGGC CAACCTCCTT CCATCTCAAG CAGATTTCAA
3B-42 TGAGGATGTG TTCCAGGGGC CAACCTCCTT CCATCTCAAG CAGATTTCAA
2B-42 TGAGGATGTG TTCCAGGGGC CAACCTCCTT CCATCTCAAG CAGATTTCAA

      201                                                     250
1B-42 CCTTTGTCAA CAGCACATGG GCTCAAAATC TAGGCTCAGG CTGGTTGGAT
3B-42 CCTTTGTCAA CAGCACATGG GCTCAAAATC TAGGCTCAGG CTGGTTGGAT
2B-42 CCTTTGTCAA CAGCACATGG GCTCAAAATC AAGGCTCAGG CTGGTTGGAT

      251                                                     300
1B-42 GACTTGCAGA TTCATGGCTG GGAGAGTGAC TCGGGCACTG CCATTTTCCT
3B-42 GACTTGCAGA TTCATGGCTG GGAGAGTGAC TCGGGCACTG CCATTTTCCT
2B-42 GACTTGAAGA TTCATGGCTG GGAGAGTGAC TCGGGCACTG CCATTTTCCT

      301                                                     350
1B-42 GAAGCCCTGG TCCAAGGGCA ACTTTAGTGA TGAGGAGATA ACTGAGCTGG
3B-42 GAAGCCCTGG TCCAAGGGCA ACTTTAGTGA TGAGGAGATA ACTGAGCTGG
2B-42 GAAGCCCTGG TCCAAGGGCA ACTTTAGTGA TGAGGAGATG ACTGAGCTGG

      351                                                     400
1B-42 TGGACCTCTT CCGAGTCTAC CTCATTGGAT TCATTCGGGA AGTGCAGGAT
3B-42 TGGACCTCTT CCGAGTCTAC CTCATTGGAT TCATTCGGGA AGTGCAGGAT
2B-42 AGGACATCTT CCGAGCCTAC TTCATTTTCT TCACTCAGGA AGTGCAGGAT

      401                        |Alpha 2 Domain              450
1B-42 CGAGTCAATG AGTTCCAGTT AGAAT|..... .......... ..GTCATAGA
3B-42 CGAGTCAATG AGTTCCAGTT AGAAT|ACCCC TTTGTGATCC AGGTCATAGA
2B-42 CGAGTCAATG AGTTCCAGTT AGAAT|ACCCC TTTGTGATCC AGGTCACAGC

      451                                                     500
1B-42 AGGCTGTGAG CTGCATTCTG GGGAGGCCAT TGAAAGCTCT TTGAGAGGAG
3B-42 AGGCTGTGAG CTGCATTCTG GGGAGGCCAT AGAAAGCTCT TTGAGAGGAG
2B-42 AGGCTGTGAG CTGCATTCTG GGGAGGCCAT AGAAAGCTCT TTGAGAGGAG

      501   B203        518
1B-42 CTTTAGGAGG ACTGGATG
3B-42 CTTTAGGAGG ACTGGATG
2B-42 CTTTAGGAGG ACTGGATG
```

Two of the three SCD1B-42 5'UT/α1 region clones (1B-42 and 3B-42) are highly homologous outwith the degenerate 5' primer site with only one single nucleotide difference at position 481, as well as a three nucleotide deletion in the 5'UT region and a seventeen nucleotide deletion at the beginning of the α2 domain (Fig. 4B). The third clone, 2B-42, contains eighteen nucleotide differences outwith the degenerate 5' primer site compared to the other two clones as well as a four nucleotide deletion in the 5'UT region. The nucleotide differences are located at positions 37, 38, 80, 89, 112, 231, 257, 351, 356, 366, 371, 377, 378, 389, 384, 387, 447 and 450. The 2B-42 clone is 99.4% identical to the known 5' end of the SCD1B-52 sequence.

3.2.5 Discussion.

The PCR technique was used to amplify the missing 5'UT/α1 regions of the truncated clones SCD1A25 and SCD1B-42. The amplification products from the reactions involving the 5' degenerate primer B199 and nested 3' GSP's were cloned and dsDNA isolated for sequencing. The inserts from five SCD1A25 5'UT/α1 region clones and three SCD1B-42 5'UT/α1 region clones were fully sequenced on each strand. The sequences within each group were aligned enabling consensus sequences to be derived for completion of the SCD1A25 and SCD1B-42 clone sequences.

Several points worthy of discussion arose during the course of this work. Thymus tissue was selected for preparation of the mRNA as the main ovine clones had been isolated from thymocyte cDNA libraries. The template cDNA for the 5'UT/α1 region PCR reactions was synthesised using random hexanucleotide primers rather than a gene specific primer or oligo (dT) primer. Use of a GSP is preferable to use of an oligo (dT) primer for isolation of 5' ends, for obvious reasons. However, it is often more advantageous to use random primed cDNA for 5' end isolation. This is because use of random primers for cDNA synthesis overcomes potential problems with inefficient binding of the GSP to the target transcript due to mRNA secondary structure. In addition, a random primed cDNA pool can be used for PCR amplification of numerous different target sequences, obviating the requirement for generation of specifically primed cDNA for each transcript of interest.

The random primed cDNA was polyadenylated for use in the RACE-PCR protocol. In theory, any nucleotide can be used in the tailing reaction. However, in practice, poly A tailing is preferred, since use of either dC or dG is more likely to lead to truncation of products due to nonspecific primer annealing during amplification. This is because homopolymers of C or G anneal with much higher affinity than do homopolymers of A or T, and 5' UT and coding sequences are naturally GC rich (Rolfs et al, 1992).

Four of the five SCD1A25 5'UT/α1 clones and two of the three SCD1B-42 5'UT/α1 clones contain small gap regions in the 5'UT domains (Fig.4A and 4B). These deletions are located within a poly A stretch of sequence which varies between seven and fifteen nucleotides in length. The location of the deletions within a poly A region in each clone suggests that they are PCR artefacts, probably produced by slippage or stuttering of the Taq pol enzyme during DNA amplification. This phenomenon of small deletions occuring in PCR products within a region of poly N (where N is A, C, G or T) sequence has also been observed during the cloning and isolation of the ovine IL-1α and β genes (C. Fiskerstrand, personal communication).

Five 5'UT/α1 region SCD1A25 clones were obtained for sequencing, three of which are identical except for two nucleotide differences from a total of 1347 nucleotides. Differences occurring within the 5' primer (B199) site were disregarded due to the degenerate nature of the primer sequence itself. Two of the three clones (1A25, 3A25 and 5A25) contain a G at position 331 and an A at position 378. Sequence 1A25 was therefore selected as the representative consensus sequence of the three clones, as it contains the appropriate nucleotides at these positions.

It is **assumed** that the two nucleotide differences between the 1A25, 3A25 and 5A25 sequences are the result of Taq polymerase infidelity. *In vitro* use of polymerase enzymes leads to misincorporation of nucleotides due to the lack of the cellular error correcting apparatus and because polymerisation conditions are often less than optimal. Cloning strategies involving the PCR are particularly error prone because the **polymerisation reaction is**

repeated many times *in vitro*. One study by Saiki et al (1988) estimated an error rate of 0.25% (or 1 in 400) following thirty cycles of amplification of a particular fragment. The estimated error frequency in a recent study of sequence variation in the gp135 gene of Maedi Visna Virus strain EV1 was 1 in 800 or 0.125% (N. Carey, PhD Thesis, University of Edinburgh, 1993). Observed error frequencies can vary more than ten fold, from 2 per 10,000 nucleotides to less than 1 per 100,000 nucleotides, depending on the target sequence and the conditions of DNA synthesis (Eckert and Kunkel, 1991). Several parameters of the PCR reaction can influence the fidelity of the polymerase enzyme, including dNTP concentrations, enzyme concentration, reaction conditions, $MgCl_2$ concentration and pH. The concept that conditions generating the greatest amount of DNA product are also optimal for fidelity of the polymerase enzyme is not necessarily true. The frequency of nucleotide differences observed for the 1A25, 3A25 and 5A25 PCR generated 5'UT/α1 regions is 1 per 675 nucleotides. This figure is comparable to the data discussed above, particularly since seventy rounds of amplification were performed prior to cloning. The two nucleotide differences detected between 1A25, 3A25 and 5A25 are both single base purine to purine transitions, one a G-A transition at position 331 and the second an A-G transition at position 378 (Fig. 4A). Ennis et al (1990) analysed the nature and frequency of PCR-derived errors produced during the amplification and cloning of full-length class I HLA cDNA. Purine to purine transitions were the most common single base substitutions detected (fourteen out of twenty-eight substitutions were purine to purine transitions).

The representative 1A25 sequence was compared to the known 5'UT/α1 region of SCD1B-52 and to the representative 3B-42 sequence, giving ~85% sequence identity in each case. This supports the view that the 1A25 sequence does not represent inadvertent amplification of the 5'UT/α1 region of either SCD1B-42 or SCD1B-52. However, it was not possible to determine the percentage identity of 1A25 compared to the 5'UT/α1 region of SCD1T10, as attempts to amplify the SCD1T10 5'UT/α1 region were unsuccessful. Consequently, the possibility that 1A25 corresponds to the 5'UT/α1 region of SCD1T10 could not be eliminated, although this possibility was minimised by careful selection of 3'GSP primer sequences. Similarly, it is possible that 1A25 corresponds to the 5'UT/α1 region of an, as yet uncloned,

ovine CD1 transcript. Given the difficulties in selecting clone specific 3' primers, it is fortuitous that three virtually identical 5'UT/α1 region clones were obtained in order to derive a consensus sequence for completion of the SCD1A25 sequence.

The two other SCD1A25 5'UT/α1 region sequences (2A25 and 4A25) are different from the representative 1A25 sequence. This is indicated by pairwise comparison of the 2A25 and 4A25 sequences to 1A25 each giving ~86% nucleotide sequence identity. Comparison of 2A25 and 4A25 sequences to SCD1B-52 each gave ~95% nucleotide sequence identity. Comparison of the 2A25 and 4A25 sequences to the representative 3B-42 5'UT/α1 region sequence (see a later paragraph) gave ~99% and 95% nucleotide sequence identity respectively. Hence it appears that the 2A25 sequence has actually been amplified from the 5'UT/α1 region of SCD1B-42. The 4A25 sequence is likely to represent inadvertent amplification of the 5'UT/α1 region of SCD1T10 or an, as yet uncloned, ovine CD1 transcript as discussed above.

Three 5'UT/α1 region SCD1B-42 clones were obtained for sequencing. Two of these clones, 1B-42 and 3B-42, are identical outwith the degenerate 5' primer site, except for a single nucleotide difference (an A-T point substitution) and two deletion sites. It is highly likely that the single nucleotide difference between 1B-42 and 3B-42 was generated during PCR amplification. The 3B-42 sequence was selected as representative of these two clones as it did not contain any gap regions. The larger of the two gap regions in 1B-42 is a splice junction deletion of seventeen nucleotides beginning at the start of the α2 domain. This deletion is highly unlikely to be a PCR artefact and may represent the use of a cryptic splice acceptor site within the α2 domain. Similarly, a human CD1 transcript has been detected in which a cryptic splice acceptor site within the transmembrane/cytoplasmic exon is utilised in preference to the site at the beginning of the exon (Woolfson and Milstein, 1994). The second deletion is three nucleotides long and located near the start of the sequence in the 5'UT region. This deletion is likely to be a PCR artefact, as discussed previously.

Comparison of 3B-42 to the 5'UT/α1 region of SCD1B-52 and the representative 1A25

sequence gave less than 96% and ~85% nucleotide sequence identities respectively. This supports the view that the 3B-42 sequence does not represent inadvertent amplification of the 5'UT/α1 region of either SCD1B-52 or SCD1A25. It was not possible to eliminate the possibility that 3B-42 corresponds to the 5'UT/α1 region of SCD1T10 or to the 5'UT/α1 regions of other as yet uncloned ovine CD1 transcripts, as discussed for the SCD1A25 5'UT/α1 regions. The 2B-42 sequence was found to be 99.5% identical to the 5'UT/α1 region of SCD1B-52 and hence appears to represent an inadvertent amplification during the PCR step. The high percentage sequence identity between the SCD1B-42 and SCD1B-52 sequences in the α2 domain (~93%- see section 3.3) rendered it particularly difficult to select 3' primer sequences which would preferentially anneal to SCD1B-42 during PCR amplification.

In view of the 5'UT/α1 region clone sequence data discussed above, the 1A25 sequence was selected for addition to the main SCD1A25 clone sequence in order to obtain full length sequence information. Similarly, the 3B-42 sequence was used to complete the sequence obtained from the main SCD1B-42 clone. The full length nucleotide sequences for both ovine CD1 molecules are presented and analysed in section 3.3.

3.3 Analysis Of The Ovine CD1 Sequences.

3.3.1 Introduction.

Four ovine CD1 clones have been isolated by library screening. Three of the clones (SCD1A25, SCD1B-42 and SCD1T10) contain inserts truncated at the 5' end. The fourth clone (SCD1B-52) contains an insert which is intact at the 5' end but has a precise deletion of the α3 domain. The missing 5' sequences for the SCD1A25 and SCD1B-42 clone inserts were obtained by PCR amplification, as detailed in section 3.2. Analyses of the SCD1A25 and SCD1B-42 sequences were therefore performed on composite sequences derived from both the main clones and from PCR 5'UT/α1 clones. This section of the thesis details the analyses performed on the ovine CD1 sequences and includes comparisons with the human, murine and rabbit CD1 sequences.

The various alignments presented in this section were generated using the PILEUP programme of UWGCG 7.3. This programme creates a multiple sequence alignment using progressive pairwise alignments, and is based on a simplification of the method of Feng and Doolittle (1987). Amino acid sequences were derived from nucleotide sequences using the TRANSLATE programme. The SCD1A25 sequence appears to contain a one nucleotide deletion at position 602 (see Fig.1) which introduces frameshifts and premature stop codons in the deduced amino acid sequence. For the purpose of this analysis, a single nucleotide (N, representing A, C, T or G) was inserted at position 602 which enabled a full-length in-frame amino acid sequence of SCD1A25 to be derived. Pairwise identity comparisons of the CD1 sequences were calculated using the GAP programme for both nucleotide and derived amino acid sequences. This programme uses the algorithm of Needleman and Munsch (1970) to find the alignment of two sequences, maximising the number of matches and minimising the number of gaps. The programme was executed using the default settings for gap weight and gap length (5.0 and 3.0 respectively for nucleotide comparisons, and 0.3 and 0.1 respectively for amino acid comparisons).

The domain boundaries for both the ovine nucleotide and amino acid sequences were

Figure 1: Aligment Of The Four Ovine CD1 Nucleotide Sequences.

The four ovine CD1 nucleotide sequences were aligned using the PILEUP programme of UWGCG 7.3.

Domain boundaries were defined by comparison to the most closely related human gene, HCD1B.

The translation initiation codon (ATG) is indicated in bold type at position 98. The translation termination codon (TGA) is indicated in bold type at position 1097.

Dashes indicate identity to the SCD1B-42 nucleotide sequence. Dots indicate gaps introduced to maintain sequence aligment.

The nucleotide numbering is based on the SCD1B-42 sequence.

The hexamer sequences underlined in the 3'UT domains denote the polyadenylation signals.

5'UT Domain

```
             1                                                    50
SCD1B-42     GAAGGAAGTC AGTACAGAGT ACTGAAAAAA AAAAAAAAAG CTGAAATTAG
SCD1A25      -G-------- -C--T----- -----...G- ----GGTTT- ----------

             51                                              97
SCD1B-42     AGATCAAATA CCAGCTCCGA GAGTCAGAAA TTCTACTTTC CAGTGAA
SCD1B-52     .......... ....------ ---------G --------C- -------
SCD1A25      ---------- ---AT--T-- ---------G C-------C- --T---G
```

Leader Domain

```
             98                                                  147
SCD1B-42     ATGCTGCTTC TACCGCTTCT GTTACTTGGA GTTATCCTCC CAGGTGGTGA
SCD1B-52     ---------- ----A----- ---------- ---------- ----------
SCD1A25      ---------- -G--A----- A--G--A-C- -----TG-G- ----------

             148
SCD1B-42     C
SCD1B-52     -
SCD1A25      -
```

Alpha 1 Domain

```
             149                                                 198
SCD1B-42     AATGAGGATG TGTTCCAGGG GCCAACCTCC TTCCATCTCA AGCAGATTTC
SCD1B-52     ---------- ---------- ---------- ---------- ----------
SCD1A25      ---------- -A-------- ---------- ------G--- TC--------

             199                                                 248
SCD1B-42     AACCTTTGTC AACAGCACAT GGGCTCAAAA TCTAGGCTCA GGCTGGTTGG
SCD1B-52     ---------- ---------- ---------- --A------- ----------
SCD1A25      G--A----C- --------T- ---------- --A------- ----------

             249                                                 298
SCD1B-42     ATGACTTGCA GATTCATGGC TGGGAGAGTG ACTCGGGCAC TGCCATTTTC
SCD1B-52     --------A- ---------- ---------- ---------- ----------
SCD1A25      -CA-T----- -C--T----- -----C---- --C-A----- -A----C---

             299                                                 348
SCD1B-42     CTGAAGCCCT GGTCCAAGGG CAACTTTAGT GATGAGGAGA TAACTGAGCT
SCD1B-52     ---------- ---------- --------C- ---------- -G--------
SCD1A25      ---------- ----T----- ------C--- ---------G -G--------

             349                                                 398
SCD1B-42     GGTGGACCTC TTCCGAGTCT ACCTCATTGG ATTCATTCGG GAAGTGCAGG
SCD1B-52     --A----A-- -------C-- --T-----TT C----C--A- ----------
SCD1A25      --A---A--A --TA------ ---------- ---T-C--TA ----------

             399                        425
SCD1B-42     ATCGAGTCAA TGAGTTCCAG TTAGAAT
SCD1B-52     ---------- ---------- -------
SCD1T10      .CG-GC-GC- G--A------ --T----
SCD1A25      -C-AT----G ---A------ C-G----
```

Alpha 2 Domain

```
           426                                                 475
SCD1B-42   ACCCCTTTGT GATCCAGGTC ATAGAAGGCT GTGAGCTGCA TTCTGGGGAG
SCD1B-52   ---------- ---------- -C--C----- ---------- ----------
SCD1T10    ----A----- ---T----G- ----C---T- ---------- -------A--
SCD1A25    ---------- --------A- ----C---T- ---------- -C-----A--

           476                                                 525
SCD1B-42   GCCATAGAAA GCTCTTTGAG AGGAGCTTTA GGAGGACTGG ATGTTTTGAG
SCD1B-52   ---------- ---------- ---T------ ---------- ------G---
SCD1T10    -------C-- ---TC----- --C--G---T -A-------- --T-CG----
SCD1A25    ---G------ ---TC----A G--------T ------T--- --T-CG----

           526                                                 575
SCD1B-42   GATCCAGAAT CATTCCTGCA TGCCTGCACC AGACAGCGGC AACAGGGGGC
SCD1B-52   ---------- --------TG C--------- ---------T -CG-------
SCD1T10    ----G----- --C--A--TG ------AG-- ----G-A--- -GTGAA-CA-
SCD1A25    C---A----- G----A--TG CA----TC-- --GAG----- -G--T--CC-

           576                                                 625
SCD1B-42   AGAAGCTTTG TGCACTCCTG AGTCAGTATC AAGGCACCTC CGATATCATT
SCD1B-52   ----TT---- -------A-- -C-------- ------T--- -------C--
SCD1T10    --TG-T---- --TTT--A-T -C------C- ------T--T G-C------A
SCD1A25    --CGTT-C-A --A----A-C -T----.-C- -T-CT-T--G T----CT--A

           626                                                 675
SCD1B-42   GAGAGACTCG TCTCAGAAAC CTGTCCTCGA TATCTCCTGG GTGTCCTCGA
SCD1B-52   ---------C ---------- ---------- -----G--A- ----------
SCD1T10    --C--G---C -----A---- ---C--C--- ---------- ----------
SCD1A25    -CT-AG---C ---T------ ---C------ ---T-----A -------T--

           676                            704
SCD1B-42   TGCAGGGAAG GCAGAACTGC AGAGGCAAG
SCD1B-52   ---------- ---------- ---------
SCD1T10    ---------- --G------- -C-------
SCD1A25    ------C--A ---------- ---------
```

Alpha 3 Domain

```
           705                                                 754
SCD1B-42   TGAAACCTGA AGCCTGGCTT TCCAGTGGCC CCACTCCTGG GCCTGGCCGC
SCD1B-52   .......... .......... .......... .......... ..........
SCD1T10    ----G----- ---------G ---------- ---------- ----------
SCD1A25    ----G----- G--------G -----C---- ---------- ----------

           755                                                 804
SCD1B-42   CTACTGCTGG TGTGCCATGT CTCAGGATTC TACCCAAAAC CTGTGCAGGT
SCD1B-52   .......... .......... .......... .......... ..........
SCD1T10    ---T------ -C-------- ---------- --T------- ----A-G---
SCD1A25    ---------- ---------- ---------- ---------- -----TG---

           805                                                 854
SCD1B-42   GATATGGATG AGGGGCAAGC AGGAGCAGCC TGGCACTCAG CAAGGAGACA
SCD1B-52   .......... .......... .......... .......... ..........
SCD1T10    ---G------ -----TG--- ---------- ---T------ ------A---
SCD1A25    ---G------ -----TG--- ----AG---- ---------- -----G----

           855                                                 904
SCD1B-42   TCATGCCCAA TGCTGACTGG ACTTGGTATC TCCGAGTAAC CCTAAATGTG
SCD1B-52   .......... .......... .......... .......... ..........
SCD1T10    ----A-T--- ------T--- ---------- ---------- ---GG-----
SCD1A25    ---------- ---AA-T--- ------C--- -G----C--- ----G-----

           905                                                 954
SCD1B-42   GCAGCTGGGG AGGCGGCTGG CCTGAGTTGC CGAGTGAAGC ACAGCAGTCT
SCD1B-52   .......... .......... .......... .......... ..........
SCD1T10    --G------- ----A----- ---------T ---------- ----------
SCD1A25    --------A- ---------- T--------- ---------- ----------

           955                            983
SCD1B-42   AGGAGACCAG GACATCATCC TCTACTGGG
SCD1B-52   .......... .......... .........
SCD1T10    ---------- --T------- -G-------
SCD1A25    ---------- ------G--- -G-------
```

TM/CYT Domain

```
          984                                              1033
SCD1B-42  GACACCCCAC ATCCATTGGC TTGATACTTG TGGCAATAAT AGTGCCCTCC
SCD1B-52  ---------- ---------- ---------- ---------- ----------
 SCD1T10  ---------T G-A------- ------T--- ---------- ----------
 SCD1A25  ---------- C----C---- C-----T--- ---------- ----T-----

          1034                                             1083
SCD1B-42  TTGATCCTTT CGATATGCCT TGCATTATGG TTTTGGAGAC GCTGGTCATA
SCD1B-52  ---------- T--------- ---------- ---------- --------C-
 SCD1T10  --------C- T------T-- ---------- --------G- ----------
 SCD1A25  C-C-----T- T---C--T-- ---------- --------G- ----------

          1084       1099
SCD1B-42  TCAGAATATC TTGTGA
SCD1B-52  --G------- ------
 SCD1T10  -----C-G-T ------
 SCD1A25  --T--C---- ------
```

3'UT Domain

```
          1100                                             1146
SCD1B-42  GCCCTGACCA TGTCTCCTTT TC.ATTTGG AATAAGTATC CAGG.AACCT
SCD1B-52  --A------G ---------- --C------ ---------- ----T-C---
 SCD1T10  T--------- ---------- C-C------ -GC--T--C- ----AGC--A
 SCD1A25  ----.AT-AC ---------- --C------ --------C- --C-A-C--A

          1147                                             1196
SCD1B-42  GAAACTTAAG TTGTCAGCCT AGGAGTCAAT CTCATTATAT TTCATCAAAT
SCD1B-52  ------C--- ---------- --------G- -----C---- ----------
 SCD1T10  ------C-G- ---A-----C ---------- T---C----- ----------
 SCD1A25  ------C--- CG-------C --T--C---- T----C---- ----------

          1197                                             1246
SCD1B-42  AATCATCACA TTTGATCAAA TCATTGTTCC TGTAGGTTGT AAGATAAATC
SCD1B-52  ---------- ---------- .--------- ---------- ------G---
 SCD1T10  --C-----T- ---------- -T-G-----T GA--A-C-.- T---------
 SCD1A25  -------TG- -CAAA----. ....------ -----A---C -----.....

          1247                                             1296
SCD1B-42  ATAATTTATA CTCTAGCAGA AACATAAATG AAAATTGTTA TTATGAGACA
SCD1B-52  ---------- -C-------- ----A----- ---------- ---------C
 SCD1T10  ---------- TA--G-TG-- --A--G-..T ---G-GTAC- ----A-----
 SCD1A25  .......... ..-AGAA-C- G-AT----G- TGTTCATA-- ----------

          1297                                             1346
SCD1B-42  ATATCACTAG CAGGATCCAC TCAGATTTCA TAGATGTGAT TTGTGAGAAA
SCD1B-52  ---------- ---------- ---------- ---------- ----------
 SCD1T10  ---------- T--A------ A-G------. .--G-----A G---------
 SCD1A25  ---CT-A -T T------TG- ---------T -G------GG A--C-G----

          1347                             1379
SCD1B-42  G.ATATACCT TGGAATAAAT GAAATGATGT ACAC...... ..........
SCD1B-52  -A-------- ---------- ---------- ----...... ..........
 SCD1T10  -A--G--T-- CA--C----- ---------- G--GGACTAA ATTGTGACAT
 SCD1A25  -A--G----- --T------- ---------- ----...... ..........

 SCD1T10  TCCTTTGGCT TCTTATTTTA TAACATTTTT TTCACTCCTA CTGAAATATC

 SCD1T10  ATTTGTCAAA ATGAGCTAAT TGTAATTATG TCAAGATAAT TGTATTTGCA

 SCD1T10  GCAGAGATGA CCAGCATTTT AACTTAATTT CATTGTATCT GCTTGGATGA

 SCD1T10  TCTATCTTGG AGAAACTGAG CTCCTTCTTC TTCCAGGGAG CCTTCCTTGG

 SCD1T10  TCCTACAGTA GAAGTAACCA TCCCTTGACT GATGTCATGC AGAGATCTAC

 SCD1T10  TTTCCTCTTT ACTCAGAACA TTGCAGTTCT TTATGTTTTC TTTTCCTTCT

 SCD1T10  TATGAGTTCT TTGAAAGAAA GTTCTTTGCA AGACTCGGCT CTAAGTGG
```

defined by comparison to the most homologous human gene. This is the HCD1B gene for each of the ovine sequences isolated (see Table 3 later in this section). The domain borders for the human, murine and rabbit sequences had been defined previously (human sequences- Martin et al, 1987; Aruffo and Seed, 1989; Balk et al, 1989; Calabi et al, 1989a: mouse sequences- Bradbury et al, 1988; Balk et al, 1991: rabbit sequences- Calabi et al, 1989b).

3.3.2 Overview Of The Ovine CD1 Nucleotide And Amino Acid Sequence Data.

By homology to the known human, murine and rabbit CD1 sequences, the ovine nucleotide sequences encode a polypeptide with three extracellular domains each approximately ninety amino acids long.

The 1377bp SCD1A25 composite cDNA sequence (Fig.1) contains a 145bp 5' untranslated region preceding the presumed translation initiation codon ATG. The 3' untranslated region is 262bp long with a consensus polyadenylation signal at position 1363-68, 15bp upstream of the poly A tail. Insertion of a single nucleotide (N) at position 602 of the sequence enabled a predicted amino acid sequence of 334 residues to be derived (Fig.2). The first sixteen amino acids constitute a signal peptide which would be cleaved to produce the mature protein.

The 1382bp SCD1B-42 composite cDNA sequence contains a 148bp 5' untranslated region followed by the translation initiation codon (Fig.1). The 3' untranslated region contains a consensus polyadenylation signal at an equivalent position to that observed in the SCD1A25 sequence. The long open reading frame extends from position 97 to position 1099 and encodes a polypeptide of 334 residues (Fig.2). Cleavage of the sixteen residue amino terminal signal sequence would produce a mature protein consisting of three extracellular domains, a hydrophobic transmembrane domain and a short cytoplasmic tail, similar to the human and murine CD1 molecules.

The 1067bp SCD1B-52 cDNA clone sequence contains a 33bp 5' untranslated region

Figure 2: Alignment Of The Deduced Ovine CD1 Amino Acid Sequences.

The four deduced ovine CD1 amino acid sequences were aligned using the PILEUP programme of UWGCG 7.3. Note that the insertion of one nucleotide at position 602 in the SCD1A25 nucleotide sequence enabled an amino acid sequence to be derived for this clone.

Domain boundaries were defined by comparison to the most closely related human sequence, HCD1B, and are denoted by vertical dashed lines.

The amino acid numbering is based on the SCD1B-42 sequence. Dashes indicate identity to the SCD1B-42 sequence. Dots indicate gaps introduced to maintain the sequence alignment.

Conserved cysteine residues are highlighted by an asterisk in bold type above the sequence.

Potential N-linked glycosylation sites (N-X-S/T, where X is any amino acid) are marked in bold type.

A plain asterisk denotes the translational stop codon at position 334.

```
           1   Leader        | Alpha 1 Domain                      50
SCD1B-42   MLLLPLLLLG VILPGG|DNED VFQGPTSFHL KQISTFVNST WAQNLGSQWL
SCD1B-52   ---------- ------|---- ---------- ---------- ----Q--G--
 SCD1T10   .......... ......|.... .......... .......... ..........
 SCD1A25   ---------A --V---|---- ---------V I-----A--- ----Q--G--

           51                                                 100
SCD1B-42   DDLQIHGWES DSGTAIFLKP WSKGNFSDEE ITELVDLFRV YLIGFIREVQ
SCD1B-52   ---K------ ---------- ------T--- M---E-I--A -F-F-TQ---
 SCD1T10   .......... .......... .......... .......... ..........
 SCD1A25   -N--LY--D- -P--T----- ---------- V---EE---- -----TL---

           101        |         *  Alpha 2 Domain              *150
SCD1B-42   DRVNEFQLEY |PFVIQVIEGC ELHSGEAIES SLRGALGGLD VLRIQNHSCM
SCD1B-52   ---------- |--------A-- ---------- ---------- -W-------A
 SCD1T10   .GLQ---F-- |-----G-A-- -----K--Q- F--AGFE--- FVS-E----V
 SCD1A25   -H-S------ |-----D-A-- ---P-K-V-- F-K--F---- FVS-K-D--A

           151             *                      *           200
SCD1B-42   PAPDSGNRGQ KLCALLSQYQ GTSDIIERLV SETCPRYLLG VLDAGKAELQ
SCD1B-52   ------T--- NF---MT--- -I-------L ---------- ----------
 SCD1T10   -E-EG-SEA- WF-VFIT--- -ILA--D--L -K-------- ---------H
 SCD1A25   -V-GG-SMA- RFYE-II--H AIC-T-AK-L L------F-S ----------

          201|   Alpha 3 Domain       *                       250
SCD1B-42   RQ|VKPEAWLS SGPTPGPGRL LLVCHVSGFY PKPVQVIWMR GKQEQPGTQQ
SCD1B-52   --|........ .......... .......... .......... ..........
 SCD1T10   --|-------- ---------- ---------- ----R-M--- -E--------
 SCD1A25   --|-------- ---------- ---------- ----W-M--- -E--E-----

           251                        *                    |  300
SCD1B-42   GDIMPNADWT WYLRVTLNVA AGEAAGLSCR VKHSSLGDQD IILYW|GHPTS
SCD1B-52   .......... .......... .......... .......... .....|-----
 SCD1T10   -N-IL----- -------D-- ---------- ---------- -----|---MY
 SCD1A25   -------N-- -H--A--D-- ---------- ---------- -V---|-----

           301  TM Domain     *       |CYT Domain 334
SCD1B-42   IGLILVAIIV PSLILSICLA LWF|WRRWSYQ NIL*
SCD1B-52   ---------- -----L---- ---|-----HR ---*
 SCD1T10   ----F----- -----L---- ---|------- TV-*
 SCD1A25   T---F----- S----L---- ---|------L T--*
```

preceding the ATG initiation codon (Fig.1). The 3' untranslated region is 262bp long and contains a consensus polyadenylation signal in the equivalent position to that observed in SCD1A25 and SCD1B-42. The 723bp open reading frame encodes a polypeptide of 241 residues (Fig.2), consisting of only two extracellular domains (α1 and α2) as well as the transmembrane and cytoplasmic domains. The α3 domain has been precisely deleted in this clone, as discussed in section 3.1.

The 1382bp SCD1T10 cDNA clone insert is a truncated CD1 sequence beginning 26bp upstream of the start of the α2 domain (Fig.1). The 3' end of the α1 domain, the α2, α3, transmembrane and cytoplasmic domains are encoded by an open reading frame of 700bp. This is followed by a long 3' untranslated region of 642bp, which lacks both a consensus polyadenylation signal and a poly A tail, as discussed in section 3.1.

3.3.3 Ovine CD1 Sequence Analysis.

A. Nucleotide Sequence Analysis.

The ovine CD1 sequences have been analysed by alignment of the sequences and by pairwise sequence comparisons. Alignment of the four ovine nucleotide sequences reveals a high degree of sequence conservation throughout all domains (Fig.1). The 3' untranslated region is the least conserved in terms of deletions or additions to the sequences, as indicated by the number of gaps required to maintain the sequence alignment. Most of the differences between the ovine nucleotide sequences are dispersed single base changes, although there are a few regions which contain clustered changes. This pattern of sequence variation has also been observed in the human CD1 sequences (Martin et al, 1986). The high degree of homology between the ovine sequences was confirmed by pairwise comparisons to determine percentage nucleotide identity. This analysis was performed for both the full-length coding sequences (Table1, plain type) and for each domain within the sequence (Table2, plain type).

Pairwise comparisons of the ovine CD1 coding sequences reveals that they are >85% identical at the nucleotide level (Table 1). This confirms the high degree of homology

Table 1: Comparisons Of The Ovine CD1 Sequences.

The table in **panel A** shows the results of pairwise comparisons of the ovine CD1 sequences. The results are presented as percentage identity values. The figures in plain type above the diagonal represent nucleotide sequence comparisons. The figures for the deduced amino acid sequence comparisons are shown in bold type below the diagonal.

Two figures are given for each SCD1B-52 comparison, since this clone lacks the sequence encoding the α3 domain.The first figure denotes comparison of sequence upstream of the α3 domain and the second figure denotes comparison of sequence downstream of the α3 domain.

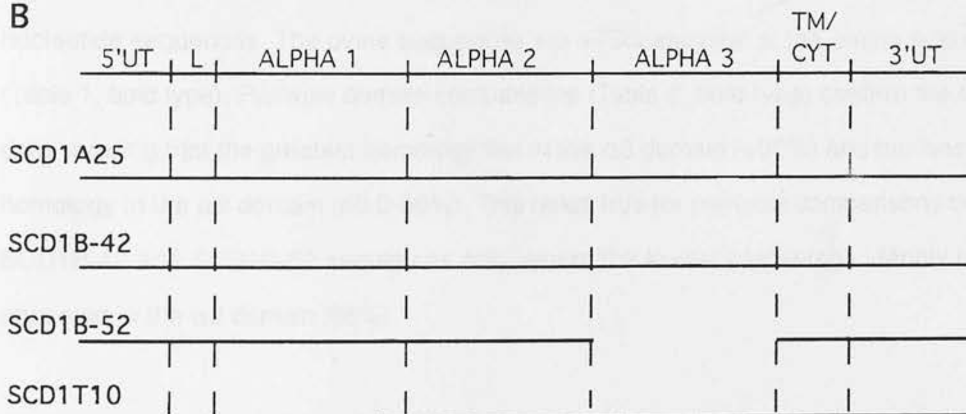
Note that the insertion of a gap at position 602 of the SCD1A25 sequence enabled an amino acid sequence to be derived for use in this analysis.

Linear representations of the four ovine CD1 sequences are shown in **panel B** to facilitate visualisation of the comparisons made in the table in panel A.

A

	SCD1A25	SCD1B-42	SCD1B-52	SCD1T10
SCD1A25		85.22	81.19/88.80	85.59
SCD1B-42	78.68		94.56/97.41	86.65
SCD1B-52	71.92/83.78	88.18/92.11		79.20/89.66
SCD1T10	78.45	79.64	66.67/81.58	

B



indicated by the sequence alignment (Fig.1). Pairwise domain comparisons of the ovine sequences shows that the greatest homology within the coding sequence is in the α3 domain (>90%) and the least homology on the α2 domain (>77%). The homology in the 3'untranslated region drops to ~65% between the SCD1A25 and SCD1T10 sequences.

There are only 36 differences between the SCD1B-42 and SCD1B-52 nucleotide sequences within the coding region, although it should be noted that SCD1B-52 completely lacks the α3 domain. The pairwise comparison of SCD1B-42 and SCD1B-52 over the same region gives a value of ~96% identity (Table 1- 96% is an average of the two figures given in the table). The nucleotide identity between equivalent domains in SCD1B-42 and SCD1B-52 is > 93% in each case (Table 2). Even the 3' untranslated regions are highly homologous (96.1% identity). A discussion of the possible allelic relationship between SCD1B-42 and SCD1B-52 is provided in section 3.5.

B. Deduced Amino Acid Sequence Analysis.

The amino acid sequences derived from the four ovine cDNA sequences have been analysed by alignment of sequences and pairwise comparisons in the same manner as the nucleotide sequences. The ovine sequences are >75% identical at the amino acid level (Table 1, bold type). Pairwise domain comparisons (Table 2, bold type) confirm the nucleotide data showing that the greatest homology lies in the α3 domain (>90%) and the least homology in the α2 domain (60.2-86%). This holds true for pairwise comparisons between SCD1B-42 and SCD1B-52 sequences only, where the lowest percentage identity is observed in the α2 domain (86%).

Alignment of the four ovine amino acid sequences (Fig.2) reveals several interesting features worthy of comment. The four cysteine residues at positions 120, 184, 224 and 279 in the α2 and α3 domains are conserved in the ovine sequences with the exception of SCD1B-52 which completely lacks the α3 domain (Fig.2). These four cysteine residues are also conserved in number and position in all other CD1 sequences and, on the basis of homology to the MHC molecules, are predicted to form intradomain disulphide bridges (Balk et al, 1989;

Table 2: Domain Comparisons Between The Ovine CD1 Sequences.

The table in **panel A** shows the results of domain by domain comparisons of the ovine CD1 sequences.The results are presented as percentage identity values.

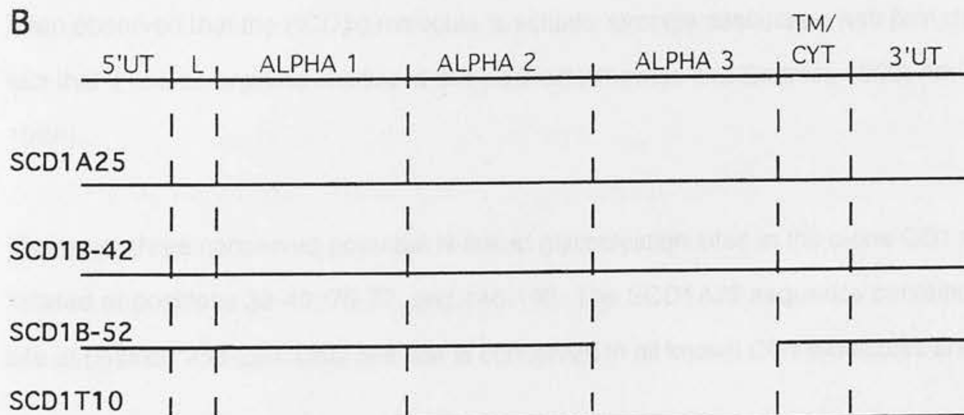
Figures in plain type represent nucleotide sequence comparisons, while figures in bold type represent deduced amino acid sequence comparisons. The insertion of one nucleotide at position 602 in the SCD1A25 nucleotide sequence enabled an amino acid sequence to be derived for this clone.

Linear representations of the four ovine CD1 sequences are shown in **panel B** to facilitate visualisation of the comparisons made in the table in panel A.

A

Region		SCD1B-42		SCD1B-52		SCD1T10	
5'UT	A25	81.92		84.85			
	B-42			93.94			
	B-52						
Leader	A25	82.35	**87.50**	84.31	**87.50**		
	B-42			98.04	**100.00**		
	B-52						
α1	A25	85.92	**81.92**	84.48	**78.72**		
	B-42			95.31	**88.30**		
	B-52						
α2	A25	76.98	**60.22**	77.34	**62.37**	79.50	**66.67**
	B-42			93.19	**86.02**	79.57	**64.52**
	B-52					79.21	**66.67**
α3	A25	91.76	**90.32**			90.32	**90.30**
	B-42					92.12	**92.47**
	B-52						
TM/CYT	A25	88.50	**85.19**	88.50	**83.78**	88.50	**83.78**
	B-42			97.35	**96.30**	90.27	**84.21**
	B-52					89.38	**81.58**
3'UT	A25	75.77		80.35		65.72	
	B-42			96.10		81.30	
	B-52					71.52	

B



Calabi et al, 1991). There are 64 amino acids in the $\alpha2$ loop and 54 amino acids in the $\alpha3$ loop of all known CD1 sequences. In ovine CD1 as in all other CD1 α chains, extra cysteine residues are observed which may mediate intermolecular associations such as covalent bonding with the CD8 molecule (Calabi and Milstein, 1986; Ledbetter et al, 1985; Snow et al, 1985). The three other cysteine residues observed in the ovine CD1 molecules are also conserved and are located at positions 149 and 163 in the $\alpha2$ domain and position 318 in the transmembrane domain. The one exception is the lack of a cysteine residue at position 163 in the SCD1A25 sequence, although there is an extra cysteine residue located at position 173 in this sequence.

The variable positions in the ovine CD1 molecules are scattered throughout the region contained within the disulphide bonded loops. This is similar to the situation observed for human CD1 and for TL/Qa class Ib molecules, but is in contrast to the situation for the classical presentation molecules, where the variable positions are mainly located outside the disulphide bonded loop (Martin et al, 1986).

Residue 280, adjacent to the second cysteine in the $\alpha3$ domain, is an arginine residue (R) and is conserved in all ovine CD1 sequences which contain an $\alpha3$ domain. An arginine residue in an equivalent position has been observed in all other human, mouse and rabbit sequences. It was initially suggested that such an arginine may contribute to the weak association of the HCD1a and HCD1b α chains with $\beta_2$microglobulin, which facilitates the exchange of human β_2m with bovine β_2m present in the culture medium (Martin et al, 1986; Calabi et al, 1991; Bernabeu et al, 1984; Amiot et al, 1986). However it has subsequently been observed that the HCD1c molecule is actually strongly associated with β_2m despite the fact that it has an arginine residue at this position (Knowles and Bodmer, 1982; Amiot et al, 1986).

There are three conserved potential N-linked glycoslyation sites in the ovine CD1 sequences, located at positions 38-40, 75-77, and 146-148. The SCD1A25 sequence contains an extra site at position 258-260. Only one site is conserved in all known CD1 molecules and this is

located at an equivalent position to the conserved N-linked glycosylation site in the MHC class II β1 domains (Balk et al, 1989; Calabi et al, 1991).

The intracytoplasmic tails of the human and murine CD1 molecules can be divided into three distinct classes (Calabi et al, 1991). The CD1a tail is short and does not contain any potential target residues for phosphorylation. The tail predicted from the CD1E nucleotide sequence is much longer than any other CD1 cytoplasmic tail. The third group comprises the HCD1b, c and d, and the MCD1.1 and MCD1.2 tail sequences. The ovine CD1 cytoplasmic tails fall into this group, as expected because of their homology to HCD1B (see section 3.3.2.3). However, only the SCD1B-42 and SCD1T10 cytoplasmic tails contain the consensus sequence S/A-Y-Q-X-I/V, where X is any amino acid, which is present in the tails of other members of this group (Balk et al, 1991a).

3.3.4 Ovine CD1 Sequences Are HCD1B Homologues.

Initial comparisons of each of the four ovine CD1 nucleotide sequences to the known human, murine and rabbit sequences revealed that all are most homologous to the human CD1B cDNA sequence (Table 3, plain type). Domain boundaries were subsequently assigned based on HCD1B homology. The derived amino acid sequences for the ovine CD1 molecules are also most homologous to HCD1b (Table 3, bold type). The second highest identity values were observed for comparisons of ovine sequences with the domestic rabbit sequence (~77-80% nucleotide identity, ~65-71.5% amino acid identity), which is itself most homologous to HCD1B. The figures for comparison of the SCD1B-52 amino acid sequence to the HCD1B and domestic rabbit sequences are 5-7% less than the values for equivalent comparisons of the other three ovine sequences. This reflects the lack of an $\alpha3$ domain in the SCD1B-52 sequence and emphasises the point that the $\alpha3$ domain is the most conserved region in the CD1 molecules (see Table 4). Comparison of the results from SCD1B-42 and SCD1B-52, which are the most closely related ovine CD1 sequences (90% amino acid identity), enables the conserved nature of the $\alpha3$ domain to be observed more clearly. By discarding the HCD1B and domestic rabbit comparisons in this data (because the ovine sequences are all homologues of these two sequences), it can be seen that the comparisons for SCD1B-52 are

Table 3: Comparisons Of The Ovine, Human, Murine and Rabbit Sequences.

The table in **panel A** shows the results of pairwise comparisons between the ovine sequences and the human, murine and rabbit sequences. The results are presented as percentage identity values.

Figures in plain type represent nucleotides sequence comparisons, while figures in bold type represent deduced amino acid comparisons. The insertion of one nucleotide at position 602 in the SCD1A25 nucleotide sequence enabled an amino acid sequence to be derived for this clone.

Two figures are given for each SCD1B-52 comparison, since this clone lacks the sequence encoding the $\alpha3$ domain. The first figure denotes comparison of sequence upstream of the $\alpha3$ domain, and the second denotes comparison of sequence downstream of the $\alpha3$ domain.

* Coding sequence for HCD1E was derived from the genomic clone rather than from a cDNA clone.

+ The domestic rabbit sequence comprised the leader and $\alpha1$ domains only (see panel B).

^ The cottontail rabbit sequence comprised the $\alpha1$, $\alpha2$ and $\alpha3$ domains only (see panel B).

Linear representations of all the CD1 molecules are presented in **panel B** to facilitate visualisation of the comparisons made in the table in panel A.

A

	SCD1A25	SCD1B-42	SCD1B-52	SCD1T10
HCD1A	70.8 **60.06**	69.12 **58.84**	63.24/63.37 **49.01/56.25**	71.78 **64.38**
HCD1B	81.02 **74.17**	79.64 **73.35**	75.23/82.76 **67.98/68.42**	79.38 **75.00**
HCD1C	70.54 **58.43**	70.17 **58.86**	65.73/68.14 **50.74/45.95**	70.94 **59.19**
HCD1D	64.24 **50.45**	64.37 **50.30**	55.68/56.90 **35.15/39.48**	68.84 **53.81**
* HCD1E	65.81 **51.70**	64.10 **51.70**	59.33/52.63 **43.43/21.88**	68.32 **56.54**
MCD1.1	60.80 **43.54**	59.78 **44.31**	55.68/45.13 **34.65/27.03**	62.89 **48.66**
MCD1.2	60.00 **42.94**	59.48 **43.71**	54.04/38.80 **32.18/27.03**	62.44 **47.77**
+ DomRab	79.27 **69.73**	79.88 **71.56**	77.44 **65.14**	
^ CtRab	65.41 **50.00**	65.33 **52.92**	55.68 **33.15**	69.71 **57.30**

B

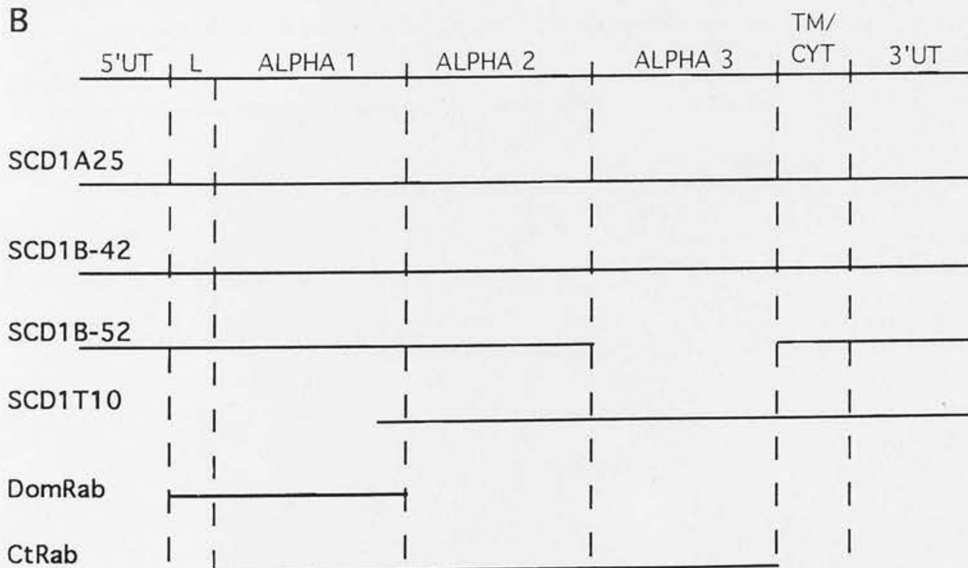


Table 4: Domain Comparisons Between The Ovine CD1 Sequences And The Most Closely Related Human Sequence, HCD1B.

The table in **panel A** shows the results of pairwise domain comparisons between the ovine CD1 sequences and the HCD1B sequence. Results are presented as percentage identity values.

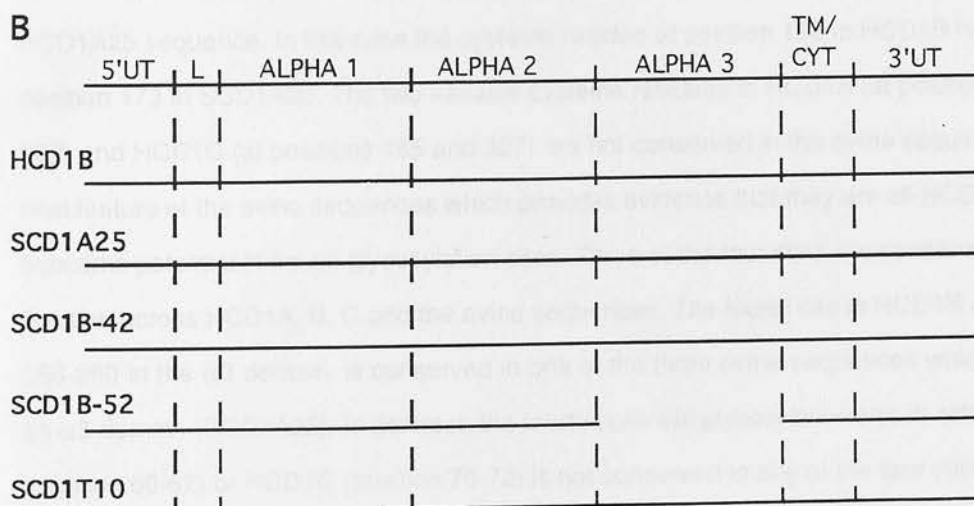
Figures in plain type represent nucleotide sequence comparisons, while figures in bold type represent deduced amino acid sequence comparisons. The insertion of one nucleotide at position 602 in the SCD1A25 nucleotide sequence enabled an amino acid sequence to be derived for this clone.

Linear representations of the four ovine CD1 sequences and the HCD1B sequence are shown in **panel B** to facilitate visualisation of the comparisons made in the table in panel A.

A

	SCD1A25		SCD1B-42		SCD1B-52		SCD1T10	
HCD1B 5'UT	73.33		77.78		75.76			
HCD1B Leader	76.47	**75.00**	66.67	**68.75**	68.63	**68.75**		
HCD1B $\alpha 1$	80.51	**72.34**	78.34	**71.28**	77.26	**69.15**		
HCD1B $\alpha 2$	77.70	**67.74**	74.91	**66.67**	74.55	**66.67**	73.84	**69.89**
HCD1B $\alpha 3$	86.74	**87.10**	86.38	**83.87**			84.95	**84.95**
HCD1B TM/CYT	77.88	**62.16**	83.19	**71.05**	82.30	**68.42**	78.76	**63.16**
HCD1B 3'UT	54.44		64.37		67.35		65.98	

B



6-20% less than the SCD1B-42 comparisons. This clearly indicates the contribution made by the more highly conserved $\alpha3$ domain to the full-length percentage identity comparisons.

The HCD1B-like nature of the ovine sequences was confirmed by aligning the amino acid sequences to each of the three "classical" human CD1 sequences, HCD1a, HCD1b and HCD1c (Fig.3). No alignments were made to CD1 sequences from the HCD1D-like group because the initial comparisons of the ovine sequences indicated that they were HCD1B-like and were therefore grouped with the "classical" CD1 molecules. The three groups of sequence alignments reveal that the ovine sequences are more homologous to HCD1B (Fig.3A) than to HCD1A (Fig.3B) or HCD1C (Fig.3C), particularly in the $\alpha1$, $\alpha2$ and transmembrane domains. The ovine sequences are identical in length to the HCD1B sequence each being 334 amino acids long (Fig.3A). In contrast, the ovine sequences are six amino acids longer than the HCD1A sequence (Fig.3B). These differences comprise two single residue additions, at positions 11 and 312, and four extra residues located in the cytoplasmic tail. Although the ovine sequences are equivalent in overall length to the HCD1C sequence, a one residue gap is observed in the ovine sequences at position 168 and a one residue gap at position 313 in the HCD1C sequence. There are seven cysteine residues in the HCD1A, B and C sequences and also in the ovine sequences. The positions of five of these are conserved across HCD1A, B and C, and the other two are variable. In the ovine sequences all seven residues are conserved in the same positions as for HCD1B, with one exception in the SCD1A25 sequence. In this case the cysteine residue at position 163 in HCD1B is found at position 173 in SCD1A25. The two variable cysteine residues in HCD1A (at positions 325 and 327) and HCD1C (at positions 155 and 327) are not conserved in the ovine sequences. The final feature of the ovine sequences which provides evidence that they are all HCD1B-like concerns potential N-linked glycosylation sites. Three of the four sites are conserved in position across HCD1A, B, C and the ovine sequences. The fourth site in HCD1B at position 258-260 in the $\alpha3$ domain, is conserved in one of the three ovine sequences which contains an $\alpha3$ domain (SCD1A25). In contrast, the fourth potential glycosylation site in either HCD1A (position 60-62) or HCD1C (position 70-72) is not conserved in any of the four ovine sequences.

Figure 3: Aligment Of The Deduced Ovine CD1 Amino Acid Sequences With The Human CD1B (FIG. 3A), Human CD1A (Fig. 3B) and Human CD1C (Fig. 3C) Amino Acid Sequences.

All sequences were aligned using the PILEUP programme of UWGCG 7.3.

Amino acid numbering is based on the top sequence in each aligment. Dashes indicate identity to the top sequence in the aligment. Dots indicate gaps introduced to maintain sequence aligment.

Cysteine residues in the top sequence of each alignment are highlighted by a bold asterisk above the sequence.

Potential N-linked glycosylation sites (N-X-S/T, where X is any nucleotide) are marked in bold type.

A plain asterisk denotes the translational stop codon.

FIGURE 3A

```
           1          Leader      |           Alpha 1 Domain            50
   HCD1B   MLLLPFQLLA VLFPGG|NSEH AFQGPTSFHV IQTSSFTNST WAQTQGSGWL
SCD1B-42   -----LL--G -IL---|DN-D V--------L K-I-T-V--- ---NL--Q--
SCD1B-52   -----LL--G -IL---|DN-D V--------L K-I-T-V--- ----------
 SCD1T10   .......... ......|.... .......... .......... ..........
 SCD1A25   -----LL--- -IV---|DN-D V--------- --I-T-A--- ---N------

           51                                                   100
   HCD1B   DDLQIHGWDS DSGTAIFLKP WSKGNFSDKE VAELEEIFRV YIFGFAREVQ
SCD1B-42   --------E- ---------- --------E- IT--VDL--- -LI--I----
SCD1B-52   ---K----E- ---------- ------T-E- MT---D---A -FIF-TQ---
 SCD1T10   .......... .......... .......... .......... ..........
 SCD1A25   -N--LY---- -P--T----- --------E- -T----L--- -LI--TL---

           101         |         *        Alpha 2 Domain        *150
   HCD1B   DFAGDFQMKY |PFEIQGIAGC ELHSGGAIVS FLRGALGGLD FLSVKNASCV
SCD1B-42   -RVNE--LE- |--V--V-E-- -----E--E- S--------- V-RIQ-H--M
SCD1B-52   -RVNE--LE- |--V--VT--- -----E--E- S--------- VWRIQ-H--A
 SCD1T10   .GLQE--FE- |--V------- -----K--Q- ---AGFE--- -V-IE-H---
 SCD1A25   -HVSE--LE- |--V--D---- ---P-K-VE- --K--F---- -V-I--D--A

           151         *                        *                200
   HCD1B   PSPEGGSRAQ KFCALIIQYQ GIMETVRILL YETCPRYLLG VLNAGKADLQ
SCD1B-42   -A-DS-N-G- -L---LS--- -TSDIIER-V S--------- --D----E--
SCD1B-52   -A-DS-T-G- N----MT--- --SDILER-- S--------- --D----E--
 SCD1T10   -E-----E-- W--VF-T--- --LAIIDR-- SK-------- --D----E-H
 SCD1A25   -V-G---M-- R-YE-----H A-CD-IAK--L-------F-S --D----E--

          201|   Alpha 3 Domain         *                       250
   HCD1B   RQ|VKPEAWLS SGPSPGPGRL QLVCHVSGFY PKPVWVMWMR GEQEQQGTQL
SCD1B-42   --|-------- ---T------ L--------- ----Q-I--- -K---P---Q
SCD1B-52   --|........ .......... .......... .......... ..........
 SCD1T10   --|-------- ---T------ L--------- ----R----- -----P---Q
 SCD1A25   --|-------- ---T------ L--------- ---------- ----EP---Q

           251                             *                 |  300
   HCD1B   GDILPNANWT WYLRATLDVA DGEAAGLSCR VKHSSLEGQD IILYW|RNPTS
SCD1B-42   ---M---D-- ----V--N-- A--------- ------GD-- -----|GH---
SCD1B-52   .......... .......... .......... .......... .....|GH---
 SCD1T10   -N-IL--D-- ----V----- A--------- ------GD-- -----|GH-MY
 SCD1A25   ---M------ -H-------- A--------- ------GD-- -V---|GH---

           301  TM Domain      *        |CYT Domain 334
   HCD1B   IGSIVLAIIV PSLLLLLCLA LWY|MRRRSYQ NIP*
SCD1B-42   --L-LV---- ---I-SI--- --F|W--W--- --L*
SCD1B-52   --L-LV---- ---I--I--- --F|W--W-HR --L*
 SCD1T10   --L-FV---- ---I--I--- --F|W--W--- TVL*
 SCD1A25   T-L-FV---- S--I--I--- --F|N--W--L T-L*
```

FIGURE 3B

```
            1   Leader    |         Alpha 1 Domain         49
   HCD1A    MLFLLLPLLA .VLPGD|GNAD GLKEPLSFHV IWIASFYNHS WKQNLVSGWL
SCD1B-42    --L-P-L--G VI---G|D-E- VFQG-T---L KQ-ST-V-ST -A---G-Q--
SCD1B-52    --L-P-L--G VI---G|D-E- VFQG-T---L KQ-ST-V-ST -A--QG----
 SCD1T10    .......... ......|.... .......... .......... ..........
 SCD1A25    --L-P-L--- VIV--G|D-E- VFQG-T---- -Q-ST-A-ST -A--QG----

            50                                                 99
   HCD1A    SDLQTHTWDS NSSTIVFLWP WSRGNFSNEE WKELETLFRI RTIRSFEGIR
SCD1B-42    D---I-G-E- D-G-AI--K- --K----D-- IT--VD---V YL-GFIREVQ
SCD1B-52    D--KI-G-E- D-G-AI--K- --K---TD-- MT---DI--A YF-FFTQEVQ
 SCD1T10    .......... .......... .......... .......... ..........
 SCD1A25    DN--LYG--- DPG-TI--K- --K----D-- VT---E---V YL-GFTLEVQ

            100       |          *  Alpha 2 Domain            149
   HCD1A    RYAHELQFEY| PFEIQVTGGC ELHSGKVSGS FLQLAYQGSD FVSFQNNSWL
SCD1B-42    DRVN-F-L--| --V---IE-- -----EAIE- S-RG-LG-L- VLRI--H-CM
SCD1B-52    DRVN-F-L--| --V----A-- -----EAIE- S-RG-LG-L- VWRI--H-CA
 SCD1T10    .GLQ-F----| --V--GIA-- ------AIQ- --RAGFE-L- ---IE-H-CV
 SCD1A25    DHVS-F-L--| --V--DIA-- ---P--AVE- --KG-SG-L- ---IK-D-CA

            150         *                     *                199
   HCD1A    PYPVAGNMAK HFCKVLNQNQ HENDITHNLL SDTCPRFILG LLDAGKAHLQ
SCD1B-42    -A-DS--RGQ KL-AL-S-Y- GTS--IER-V -E----YL-- V------E--
SCD1B-52    -A-DS-TRGQ N--ALMT-Y- GIS--LER-- -E----YL-- V------E--
 SCD1T10    -E-EG-SE-Q W--VFIT-Y- GILA-IDR-- -K----YL-- V------E-H
 SCD1A25    -V-GG-S--Q R-YELII-YH AIC-TIAK-- LE----YFLS V------E--

           200|  Alpha 3 Domain      *                         249
   HCD1A    RQ|VKPEAWLS HGPSPGPGHL QLVCHVSGFY PKPVWVMWMR GEQEQQGTQR
SCD1B-42    --|-------- S--T----R- L--------- ----Q-I--- -K---P---Q
SCD1B-52    --|........ .......... .......... .......... ..........
 SCD1T10    --|-------- S--T----R- L--------- ----R----- -----P---Q
 SCD1A25    --|-------- S--T----R- L--------- ---------- ----EP---Q

            250                            *                | 299
   HCD1A    GDILPSADGT WYLRATLEVA AGEAADLSCR VKHSSLEGQD IVLYW|EHHSS
SCD1B-42    ---M-N--W- ----V--N-- -----G---- ------GD-- -I---|G-PT-
SCD1B-52    .......... .......... .......... .......... .....|G-PT-
 SCD1T10    -N-ILN--W- ----V--D-- -----G---- ------GD-- -I---|G-PMY
 SCD1A25    ---M-N-NW- -H-----D-- -----G---- ------GD-- -----|G-PT-

                                           CYT Domain
            300      TM Domain        |    * * 328
   HCD1A    VGFIILAVIV P.LLLLIGLA LWF|RKRCFC* ....
SCD1B-42    I-L-LV-I-- -S-I-S-C-- ---|WR-WSYQ NIL*
SCD1B-52    I-L-LV-I-- -S-I---C-- ---|WR-WSHR NIL*
 SCD1T10    I-L-FV-I-- -S-I---C-- ---|WR-WSYQ TVL*
 SCD1A25    T-L-FV-I-- SS-I---C-- ---|WR-WSYL TIL*
```

FIGURE 3C

```
           1          Leader   |         Alpha 1 Domain             50
    HCD1C  MLFLQFLLLA LLLPGG|DNAD ASQEHVSFHV IQIFSFVNQS WARGQGSGWL
 SCD1B-42  --L-PL---G VI---G|--E- VF-GPT---L K--ST---ST --QNL--Q--
 SCD1B-52  --L-PL---G VI---G|--E- VF-GPT---L K--ST---ST --QN------
  SCD1T10  .......... ......|.... .......... .......... ..........
  SCD1A25  --L-PL---- VIV---|--E- VF-GPT---- ---ST-A-ST --QN------

           51                                                  100
    HCD1C  DELQTHGWDS ESGTIIFLHN WSKGNFSNEE LSDLELLFRF YLFGLTREIQ
 SCD1B-42  -D--I---E- D---A---KP -------D-- ITE-VD---V --I-FI--V-
 SCD1B-52  -D-KI---E- D---A---KP ------TD-- MTE--DI--A -FIFF-Q-V-
  SCD1T10  .......... .......... .......... .......... ..........
  SCD1A25  -N--LY---- DP--T---KP -------D-- VTE--E---V --I-F-L-V-

           101        |         *    Alpha 2 Domain             150
    HCD1C  DHASQDYSKY| PFEVQVKAGC ELHSGKSPEG FFQVAFNGLD LLSFQNTTWV
 SCD1B-42  -RVNEFQLE-| --VI--IE-- -----EAI-S SLRG-LG--- V-RI--HSCM
 SCD1B-52  -RVNEFQLE-| --VI--T--- -----EAI-S SLRG-LG--- VWRI--HSCA
  SCD1T10  .GLQEFQFE-| --VI-GI--- ------AIQS -LRAG-E--- FV-IE-HSC-
  SCD1A25  --V-EFQLE-| --VI-DI--- ---P--AV-S -LKG--G--- FV-IK-DSCA

           151 *      *                       *                200
    HC1DC  PSPGCGSLAQ SVCHLLNHQY EGVTETVYNL IRSTCPRFLL GLLDAGKMYV
 SCD1B-42  -A-DS-NRG- KL-A--S.-- Q-TSDIIER- VSE----Y-- -V-----AEL
 SCD1B-52  -A-DS-TRG- NF-A-MT.-- Q-ISDILER- LSE----Y-- -V-----AEL
  SCD1T10  -E-EG--E-- WF-VFIT.-- Q-ILAIIDR- LSK----Y-- -V-----AEL
  SCD1A25  -V--G--M-- RFYE-II.-- HAICD-IAK- LLE----YF- SV-----AEL

           201|   Alpha 3 Domain        *                          250
    HCD1C  HRQ|VRPEAWL SSRPSLGSGQ LLLVCHASGF YPKPVWVTWM RNEQEQLGTK
 SCD1B-42  Q--|-K----- --G-TP-P-R ------V--- -----Q-I-- -GK---P--Q
 SCD1B-52  Q--|....... .......... .......... .......... ..........
  SCD1T10  ---|-K----- --G-TP-P-R ------V--- -----R-M-- -G----P--Q
  SCD1A25  Q--|-K----- --G-TP-P-R ------V--- -------M-- -G---EP--Q

           251                              *              | 300
    HCD1C  HGDILPNADG TWYLQVILEV ASEEPAGLSC RVRHSSLGGQ DIILYW|GHHS
 SCD1B-42  Q---M----W ----R-T-N- -AG-A----- --K-----D- ------|--PT
 SCD1B-52  .......... .......... .......... .......... ......|--PT
  SCD1T10  Q-N-IL---W ----R-T-D- -AG-A----- --K-----D- ------|--PM
  SCD1A25  Q---M---NW --H-RAT-D- -AG-A----- --K-----D- --V---|--PT

                                           CYT Domain
           301       TM Domain        |  *       334
    HCD1C  SMNWIALVVI VP.LVILIVL VLWF|KKHCSY QDIL*
 SCD1B-42  -IGL-LVAI- --S-ILS-C- A---|WRRW-- -N--*
 SCD1B-52  -IGL-LVAI- --S-IL--C- A---|WRRW-H RN--*
  SCD1T10  YIGL-FVAI- --S-IL--C- A---|WRRW-- -TV-*
  SCD1A25  -TGL-FVAI- -SS-IL--C- A---|WRRW-- LT--*
```

Altogether, the features highlighted above confirm the results from the percentage identity comparisons that all four ovine CD1 sequences are most homologous to the human CD1B sequence.

3.3.5 Do SCD1B-42 and SCD1B-52 Represent Alleles From A Single Locus?

The SCD1B-42 and SCD1B-52 clones were both isolated from the ST-2 thymocyte cDNA library which had been generated from one eight week old lamb thymus (Hein et al, 1989). Despite the lack of the $\alpha3$ domain in SCD1B-52, comparison of these sequences at the nucleotide level gives a value of ~96% identity overall (Table 1) and >93% identity in each domain including the 5' and 3' untranslated regions (Table 2). These results indicate that SCD1B-42 and SCD1B-52 may represent alleles of a single gene. An alternative possibility is that SCD1B-42 and SCD1B-52 represent the result of a recent gene duplication event in the ovine CD1 locus. All other pairwise nucleotide comparisons between the ovine CD1 sequences give values of 85-87% identity overall and values of 81.3 % or less in the 3' untranslated regions. Similar results were obtained for amino acid sequence comparisons- the overall identity between SCD1B-42 and SCD1B-52 is ~90% in contrast to ~75-80% for the other comparisons (Table 1). Domain by domain comparisons between SCD1B-42 and SCD1B-52 coding sequences gave values of >86% in each case (Table 2).

It is particularly notable that the SCD1B-42 and SCD1B-52 sequences are ~96% identical in the 3' untranslated regions (Table 2), in contrast to <81.3% identity for comparisons between other ovine CD1 sequences in this region. This observation is rather surprising given that 3' untranslated regions lack the structural constraints of coding sequences and therefore retain less percentage sequence identity than coding domains. Indeed, the human 3' untranslated sequences are 40-60% identical and the two mouse 3' untranslated sequences are ~55% identical. In human and murine polymorphic MHC class I genes the 3' untranslated regions of alleles from the same locus show a mean homology of 96+/-2.3%, whereas the homology between alleles from different loci is in the range 74-95% (Ennis et al, 1988). Similar results were obtained for pairwise comparisons of the three extracellular domains. An analysis of human class I sequences has shown that there can be up to 11% variation across the 274

amino acids encoding the $\alpha 1$, $\alpha 2$ and $\alpha 3$ domains of alleles from a single locus, and that the majority of these differences are in the $\alpha 1$ and $\alpha 2$ domains (Bjorkman and Parham, 1990). Comparison of SCD1B-42 and SCD1B-52 in the equivalent regions, i.e. across the $\alpha 1$ and $\alpha 2$ domains only, gives a percentage identity value of ~87%. This value is comparable to the greatest differences observed between MHC class I alleles, given that the homology of the $\alpha 3$ domains of SCD1B-42 and SCD1B-52 is likely to be even greater than the ~90% identity between other ovine $\alpha 3$ domains. Given that it is probably more appropriate to compare the CD1 genes to other nonpolymorphic class Ib genes, it is interesting to note that alleles of the class Ib genes in mice differ in sequence by 1% or less, in contrast to the human classical class I alleles (Hedrick, 1992). Significant polymorphism of CD1 sequences would, however, be unprecedented. Previous studies in mouse and man have shown that CD1 polymorphism is extremely limited (Calabi et al, 1991). In humans, only minor differences, mainly single base substitutions, have been detected in the coding regions of cDNA clones isolated from several individuals, and it is likely that these were cloning or sequencing artefacts. In the mouse, only one RFLP has been detected by Southern blot analysis of 11 inbred laboratory mouse strains using an $\alpha 3$ region probe (A. Bradbury, PhD Thesis). A more sensitive analysis was conducted by nuclease protection assays using exon specific probes (Bradbury et al, 1988). This analysis was expected to detect most multiple mismatches and some but not all single base differences within the probe length. Two independent variants of the $\alpha 3$ domain were detected, probably involving single base changes. The extent of CD1 polymorphism in the rat was similarly assessed by Southern blot analysis in 12 strains of laboratory rat using 5'UT/a1/a2/a3 and TM/CYT/3'UT probes (Ichimiya et al, 1994). Three allelic variants of the rat CD1 gene were detected using the TM/CYT/3'UT probe. However, the nature of this probe did not enable variation in the TM/CYT domain to be distinguished from variation in the 3'UT region.

The question of an allelic relationship between SCD1B-42 and SCD1B-52 can only be resolved when the genomic organisation of the ovine CD1 gene family has been established. It is particularly difficult to resolve the question of allelic relationships in animals like sheep due to the unavailability of inbred strains. This problem does not arise in inbred laboratory mice

where a non-allelic relationship can be confirmed by the presence of both gene products in an inbred strain. For example, the putative nonclassical allelic gene products Qa-1[a] and Qa-1[b] were investigated using PCR. The results indicated that B10.M mice express Qa-1[b] but not Qa-1[a], suggesting that these molecules are the products of a single polymorphic gene (Connolly et al, 1993). It is possible that SCD1B-42 and SCD1B-52 represent genes produced by a very recent duplication event within the CD1 gene complex, in a similar manner to that proposed for the two mouse genes (Hughes, 1991). This could be investigated by hybridisation analysis of genomic DNA using probes which would differentiate between the two sequences. Given the small number of differences between SCD1B-42 and SCD1B-52 it may be difficult to generate such a probe, although one promising region is the section between nucleotides 349 and 389, within which there are nine differences between the two sequences. An alternative approach would be to analyse genomic DNA digested with a regular cutter enzyme from a variety of sheep using a probe which does not differentiate between the two sequences. If SCD1B-42 and SCD1B-52 represent two different genes, two distinct bands would always be observed, even in homozygous sheep.

3.3.6 Relationship Between Ovine CD1 and MHC Class I Molecules.

The resolution of the 3-D crystal structure of MHC class I molecules has permitted not only the identification of structural features important for peptide binding and T cell recognition but also characterisation of residues involved in interaction with β2m and CD8 as well as peptide docking residues. The following three sections deal with the analysis of ovine CD1 molecules for the potential to interact with β2m, CD8 and peptide ligands.

3.3.6.1 Interaction of Sheep CD1 α Chains With β2-microglobulin.

Balk and colleagues (1989) analysed the β2m contact residues in human CD1 heavy chains using the unrefined crystal structure data of HLA-A2 at 0.35nm resolution (Bjorkman et al, 1987). They found that none of the possible contact points between the HLA class I α3 domain and β2m were conserved in CD1 with the exception of an Asp residue at position 238 in HLA class I. Similarly, Tysoe-Calnon and colleagues (1991) used the refined crystallographic data for HLA-A2 and HLA-Aw68.1 at 0.26nm resolution (Garrett et al, 1989) in conjunction

with sequence alignments of ninety classical and nonclassical α chains and secondary structure predictions to analyse class I/β2m interactions. The human CD1 amino acid sequences were included in this analysis. The majority of conserved β2m contact residues in CD1 were located in the α2 domain and the major contact region in the α3 domain. Together with the results of secondary structure predictions, they concluded that the top part of the α3 domain would bind to β2m in an MHC class I-like manner, but that the α1 and α2 domains may fold differently and hence bind to β2m in a non-MHC class I-like manner. The authors point out that biochemical evidence exists to support this conclusion in that CD1 has been shown to form an intermolecular complex with class I on human thymocytes (Amiot et al, 1988). If the CD1 α chain is binding to β2m already bound to the class I α chain, it is presumed that the β2m contact sites cannot be the same in both α chains.

The HLA-A2 crystal structure has since been further refined by Saper et al (1991) and has been used to analyse the potential β2m contact residues in the ovine CD1 α chains (Figure 4). The most homologous human sequence, HCD1B, has also been included in this analysis. The results show that of the thirty-one β2m contact residues identified in the refined HLA-A2 crystal structure, the majority are neither conserved nor conservatively replaced in the ovine CD1 sequences (Table 5). Three of the nine contact sites in the α2 domain are completely conserved between HLA-A2, HCD1B and the ovine CD1 sequences. The alanine (A) residue at position 117 in the α2 domain of HLA-A2 is conserved in three of the four ovine sequences and in HCD1B. The fourth ovine sequence has a glycine (G) residue at this position, identical to that observed in the ovine class I sequence. The α3 domain of the class I α chain is the main site of interaction with β2m. Of the nine contact residues located in the α3 domain loop that interact with β2m (residues 231-244), only two are conserved in two of the three ovine sequences- the proline (P) at position 234 and the aspartic acid (D) at position 237. The proline residue at position 234 is also conserved in the HCD1B sequence. It is also notable that of the thirty-one β2m contact residues identified in the HLA-A2 structure, there are only six which differ between sheep and human class I molecules (at positions 9, 23, 116, 117, 121 and 235- see Table 5). Three of these differences are conservative substitutions- leucine/isoleucine at position 23, glycine/alanine at position 117, and arginine/lysine at

Figure 4: Alignment Of The Deduced Ovine CD1 Amino Acid Sequences With Other MHC and CD1 Sequences.

The deduced ovine CD1 amino acid sequences have been aligned with the human class I consensus sequence (Tysoe-Calnon et al, 1991), the mouse class I consensus, mouse class II consensus and human class II consensus sequences (Martin et al, 1986) and the mouse CD1 consensus sequence (Bradbury et al, 1988). The alignment is based on that of Tysoe-Calnon et al, 1991.

Dashes indicate positions where no consensus residue was available, as defined in the original papers.

Blank spaces are gaps introduced to maintain the sequence alignment.

The numbered vertical dashes mark the end of every block of ten amino acids in the human class I consensus sequence.

Residues marked in bold type are amino acids in human class I molecules which interact with $\beta_2$microglobulin (Saper et al, 1991).

Underlined residues are amino acids in human class I molecules which interact with CD8 (Salter et al, 1989).

Shadowed residues are amino acids which have been identified as being essential for anchoring the amino and carboxy terminals of peptides in the peptide binding groove of class I molecules (Madden et al, 1991).

Alpha 1 Domain

```
                                  10        20        30        40         50        60        70        80        90
                                   |         |         |         |          |         |         |         |         |
H class I consensus          GSHSMRYFYTSVSRPGRGEPRFIAVGYVDDTQFVRFDSDAASPR MEPRAPWIEQEGPEYWDRETQIVKAQSQTDRESLRTLRGYYNQSEA
SCD1B-42   DNEDVFQGPTSFHLKQISTFVNSTWAQNLG   SQWLDDLQIHGWESDSGTAIFLKPWSKGNFSDE EITELVDLFRVYLIGFIR  EVQDRVNEFQLEY
SCD1B-52   DNEDVFQGPTSFHLKQISTFVNSTWAQNQG   SGWLDDLKIHGWESDSGTAIFLKPWSKGNFTDE EMTELEDIFRAYFIFFTQ  EVQDRVNEFQLEY
SCD1A25    DNEDVFQGPTSFHVIQISTFANSTWAQNQG   SGWLDNLQLYGWDSDPGTTIFLKPWSKGNFSDE EVTELEELFRVYLIGFTL  EVQDHVSEFQLEY
```

Alpha 2 Domain

```
                              100       110       120       130       140       150          160       170       180
                               |         |         |         |         |         |            |         |         |
H class I consensus   GSHTLQRMYGCDVGPDGRLLRGYHQYAYDGKDYIALNEDLRSWTADTAAQITQRKWEAA    RVAEQLRAYLEGTCVEWLRRYLENGKETLQRA
SCD1B-42               PFVIQVIEGCELHS GEAIESSLRGALGGLDVLRIQ  NHSCMPAPDSGNRGQKLCALLSQYQGTSDIIERLVSETCPRYLLGVLDAGKAELQRQ
SCD1B-52               PFVIQVIAGCELHS GEAIESSLRGALGGLDVWRIQ  NHSCAPAPDSGTRGQNFCALMTQYQGISDIIERLLSETCPRYLLGVLDAGKAELQRQ
SCD1T10                PFVIQGIAGCELHS GKAIQSFLRAGFEGLDFVSIE  NHSCVPEPEGGSEAQWFCVFITQYQGILAIIDRLLSKTCPRYLLGVLDAGKAELHRQ
SCD1A25                PFVIQDIAGCELHP GKAVESFLKGAFGGLDFVSIK  NDSCAPVPGGGSMAQRFYELIIQYHAICDTIAKLLLETCPRYFLSVLDAGKAELQRQ
```

Alpha 3 Domain

```
                             190          200       210       220         230       240         250       260        270
                              |            |         |         |           |         |           |         |          |
H class I consensus    DPPKTHVTHHP ISDHEA TLRCWALGFYPAEITLTWQRD GEDQTQDTELVETRPAGDGTFQKWAAVVV PSGEEQRYTCHVQHEGLP KPLTLRW
M class I consensus    DSPKAHVT-H-R-- --V TLRCWALGFYPADITLTWQLN GEELTQDMELVETRPAGDGTFQKWASVVV PLGKEQ-YTC-V-H-GLPEP LTLRW
HCD1 consensus         VKPEAWLS-GPSPGPG-L -LVCHVSGFYPKPVWV-WMR  GEQEQ-GTQ-GDILPNAD-TWYLRATL-VA- GEAAGLSCRVKHSSL-GQDI-LYW
MCD1 consensus         EKPVAWLSSVPSSAHGH- QLVCHVSGFYPKPVWVMWMR  GDQEQQGTHRGDFLPNADETWYLQATLDVEA GEEAGLACRVKHSSLGGQDIILYW
SCD1 consensus         VKPEAWLSSGPTPGPGRL LLVCHVSGFYPKPV-V-WMR  G-QEQPGTQQG-I--NADWTW-LR-TL-VAA GEAAGLSCRVKHSSLGDQDI-LYW
H class II consensus   --P-V---PS----L-HHNLLVC-V--FYP--I-VRWF-N-QEE--G-VST-LI- NGDWTF--LVMLE--P--G- VY-C-VEH-S--SP -TVEW
M class II consensus   --P-V------T--L-HHN-LVCSV-DFYP--I-VRWFRNG-EE--G--ST-L-R NGDWTFQ-LVMLE--P--GE VYTC-VEHPSL--P VTVEW
```

Table 5: Analysis of Known Class I /β2m Interaction Sites in The Ovine CD1 Molecules.

This analysis was based on the crystal structure results of Saper at al, 1991.

The position numbers are for the HLA consensus sequence shown in Figure 4.

The sheep class I sequence was derived from the nucleotide sequence of pSC112, a full-length class I clone obtained by Grossberger and coworkers, 1990 (Acc. No. M34676).

Residues in bold type are identical to those in the same position in both the sheep class I sequence and the HLA consensus sequence.

Residues with a single underline are identical to that observed in the same position in the HLA class I consensus sequence but not to that in the sheep class I sequence.

Residues with a double underline are identical to that observed in the same position in the sheep class I sequence but not to that in the HLA consensus sequence.

The * sign indicates that due to the introduction of gaps in the CD1 sequence which are required to maintain alignments (see Fig. 4), no equivalent residue to that in the class I sequence exists in the CD1 sequences.

Domain	Position	Sheep Class I	HLA Class I Consensus	SCD1	HCD1B
alpha 1	8	F	F	3S	S
	9	S	Y	3T	S
	10	T	T	3F	F
	12	V	V	3N	N
	23	L	I	*	*
	25	V	V	3S	S
	27	Y	Y	3W	W
	32	Q	Q	**2Q**,1K	**Q**
	35	R	R	3G	G
	48	R	R	3W	W
alpha 2	94	T	T	4V	E
	96	Q	Q	**4Q**	**Q**
	115	Q	Q	3R,1K	R
	116	F	Y	3G,1A	G
	117	G	A	3A,1G	A
	119	D	D	3G,1E	G
	120	G	G	**4G**	**G**
	121	R	K	4L	L
	122	D	D	**4D**	**D**
alpha 3	192	H	H	3G	G
	202	R	R	3V	V
	204	W	W	3H	H
	231	E	E	2D,1N	D
	232	T	T	3I	I
	235	P	P	**2P**,1L	**P**
	236	S	A	3N	N
	237	G	G	3A	A
	238	D	D	**2D**,1N	N
	239	G	G	3W	W
	242	Q	Q	2Y,1H	Y
	244	W	W	3R	R

position 235.

In contrast to the results obtained for the analysis of CD1 α chain/β2m interactions, Flajnik and colleagues (1993) found that the majority of the β2m interaction sites in Xenopus classical class I molecules (defined on the basis of the refined HLA-A2 crystal structure of Saper and colleagues, 1991) are identical or conservatively replaced in Xenopus nonclassical class Ib molecules. This suggests that both classical and nonclassical Xenopus class I genes encode α chains capable of associating with a putative Xenopus β2m molecule. Given that the Xenopus class Ib genes are non MHC-linked (Flajnik et al, 1993), it is interesting that their products have the potential to interact with β2m in a classical class I-like manner in contrast to the products of the non MHC-linked CD1 genes.

The data suggest that the ovine CD1 genes encode heavy chains which interact with β2m in a manner which differs from the known MHC classI/β2m interaction but which is similar to the HCD1B/β2m interaction. Determination of the ovine CD1 crystal structure would enable the exact nature of the α chain/ β2m interaction to be defined.

3.3.6.2 Interaction of Sheep CD1 α Chains With CD8.

Characterisation of mutant murine and human class I molecules with substitutions at positions 227 and 245 indicated that the α3 domain is the principal site of contact with the CD8 molecule (Potter et al, 1987; Salter et al, 1989). Salter and colleagues (1990) substantiated this hypothesis and defined a minimum CD8 binding site by measuring levels of adhesion between CD8 and forty-eight point mutants of HLA-A2.1. Three clusters of α3 residues contribute to the binding with the CD8 α chain- residues 223-229, 233-235 and 245-247. The exposed negatively charged loop (residues 223-229) was found to play a dominant role in adhesion to CD8 with Q226 and D227 thought to be the most important residues. The mutants were also analysed for T cell recognition and the data showed a good correlation between CD8 binding and cytotoxic T cell recognition.

The three clusters of residues comprising the putative CD8 binding site have been examined

in the $\alpha3$ domain of the CD1 sequences (Figure 4). Only one of the residues in the dominant loop of HLA-A2.1 is conserved in the CD1 sequences (T228). The absence of residues Q226 and D227 in this region is particularly noteworthy. Two of the three ovine sequences contain a P residue identical to that at position 235 (equivalent to position 255 in Figure 2) in the second CD8 binding cluster. Similarly, only one of the three ovine sequences contains an A residue at position 245 (equivalent to position 265 in Figure 2) identical to that in the third CD8 binding cluster of HLA-A2.1. The most homologous human molecule, HCD1B, contains four of the thirteen residues implicated in interaction with CD8 (Q226, T228, P235, A244, which are in equivalent positions to Q245, T248, P255 and A264 in Figure 2).

Similar studies of potential CD8 interaction sites have been carried out in a variety of other species including shark, chicken and frog. The shark class I sequence was found to contain a Q residue at an equivalent position to the major CD8 contact residue Q226 in HLA-A2.1 (Hashimoto et al, 1992). One of the three residues comprising the second cluster is also conserved in the shark sequence (P235). In the chicken class I sequence both Q226 and D227, the most critical CD8 binding residues, are conserved (Kaufman et al, 1992), although none of the other residues that have been described as important for CD8 binding are conserved. In the frog, neither of these two residues are conserved in the Xenopus classical class I sequence nor in the Xenopus nonclassical class Ib sequences (Shum et al, 1993; Flajnik et al, 1991). The authors suggest that compensatory changes will probably be detected in the MHC binding site of the as yet uncloned Xenopus CD8 molecule. Interestingly, Bahram and coworkers (1994) have recently reported the identification of a family of related sequences in the human MHC which are only distantly related to the MHC classical class I chains. These MIC genes (MHC class I chain-related genes) contain none of the residues implicated in binding of CD8 to class I molecules.

This analysis suggests that the ovine CD1 α chains do not contain typical CD8 binding residues, although it is possible that other residues outwith the minimum CD8 binding site are able to perform this function.

3.3.6.3 Interaction of Sheep CD1 With Peptide.

The crystallographic structure of the HLA-A2 molecule has provided a structural basis for the direct binding of peptide to class I molecules (Bjorkman et al, 1987; Bjorkman and Parham, 1990). A putative binding groove is formed by the $\alpha1$ and $\alpha2$ domains of the class I heavy chain. The sides of the groove are formed by two long α-helices and the base of the groove by an eight-stranded antiparallel β-pleated sheet. The groove is located on the top surface of the molecule and was found to be filled with electron dense material not accounted for by the HLA molecule. This material was proposed to be peptide(s) filling the binding site. Refinement of the HLA-A2 structure and solution of the structure for HLA-Aw68 yielded further insights into the detailed architecture of the peptide binding groove (Garrett et al, 1989). The structural differences between the two molecules demonstrate how polymorphism creates and alters subpockets (A to F) in the peptide groove. Determination of the X-ray crystallographic structure of HLA-B27 revealed electron density in the antigen binding grove which was interpreted to be the image of a nonameric peptide in a largely extended conformation (Madden et al, 1991; Madden et al, 1992). The clarity of the electron density in the groove enabled detailed interactions of the peptide with the class I molecule to be observed. The crystallographic structures of various murine and human class I molecules complexed with peptides of different lengths have now been determined (Fremont et al, 1992; Matsumara et al, 1992; Zhang et al, 1992). The overall structure of each molecule is similar but the important feature of all is the conservation of non-polymorphic residues at each end of the peptide binding cleft. These residues form extensive hydrogen bonds with the peptide amino and carboxy termini which determine the orientation of the peptide in the cleft and restict the length of peptides which can be accommodated (van Bleek et al, 1993; Forquet et al, 1993, Parham, 1992). Slightly longer peptides are able to bind by a bulging out of the middle portion of the peptide to compensate for the extra length.

Residues which are conserved in essentially all of the human classical class I proteins have been identified that are crucial in anchoring the amino and carboxy termini of antigenic peptides (Madden et al, 1991; Madden et al, 1992). All eight of these docking residues are conserved in the known ovine class I sequences (Grossberger et al, 1990), but none are

conserved in the ovine CD1 sequences (Table 6 and Figure 4), indicating that if the CD1 molecules can bind peptides, as suggested by the CD1 antigen presentation study by Porcelli and colleagues (1992), then they do so in an alternative manner to that observed for classical class I molecules. Determination of the crystal structure of CD1 together with attempts to elute and analyse peptides from CD1 molecules should help clarify the nature of putative CD1 antigen presentation.

Similar studies have been carried out to search for peptide docking residues in other nonclassical class I molecules. The protein product of the H2-M3 gene in mice has been shown to present N-formyl peptides of mitochondrial and prokaryotic origin to $CD8^+$ T cells (Shawar et al, 1991; Vyas et al, 1992). Two of the four carboxy terminal docking residues are conserved in the M3 molecule (Tyr84 and Thr143) but Lys146 is replaced by Arg and Trp147 with Leu (Wang et al, 1991). At the other end of the cleft, three of the four conserved Tyr residues which interact with the peptide amino terminus are present in M3, but Tyr171 is replaced by a Phe residue. The extra space formed by the replacement of tyrosine 171 with phenylalanine could accommodate an N-formylated peptide terminus. In the Xenopus classical class I molecule, all eight docking residues are conserved (Flajnik et al, 1993). The situation for the Xenopus XNC (nonclassical class I) molecules is somewhat different. There is a lack of conserved residues at the peptide carboxy terminus site although there is better conservation of amino terminus docking residues (Flajnik et al, 1993). Mutational analysis has recently suggested that amino terminal docking is more critical for peptide binding than carboxy terminal binding (Latron et al, 1992). In the murine nonclassical class I molecule Qa-1^a, the four residues which stabilise the amino terminus of the peptide are all present and two of the four carboxy terminal docking residues are also conserved (Connolly et al, 1993), suggesting that Qa-1^a is capable of binding peptides in a similar manner to classical class I molecules. The Thr143 and Trp147 are both replaced by Ser in the Qa-1^a molecule. The effect of these changes is to open up the binding groove, perhaps permitting the binding of peptides with bulky carboxy termini. In the molecules encoded by the MIC genes, a recently identified second lineage of conserved MHC genes distantly related to the class I genes, only

Table 6: Analysis Of Known Class I Peptide Docking Residues In The Ovine CD1 Molecules.

The docking residue details are from Madden et al, 1991.

The residue numbering is based on the HLA consensus sequence shown in Figure 4.

Docking Residues	SCD1 Sequences
Peptide N-terminus	
Tyr 7 (Y)	3 Ile (I)
Tyr 59 (Y)	1 Ile (I), 1 Met (M), 1 Val (V)
Tyr 159 (Y)	4 Leu (L)
Tyr 171 (Y)	4 Val (V)
Peptide C-terminus	
Tyr 84 (Y) or Arg 84 (R)	2 Asn (N), 1 Ser (S)
Thr 143 (T)	2 Gly (G), 2 Ala (A)
Lys 146 (K)	1 Leu (L), 3 Phe (F)
Trp 147 (W)	3 Cys (C), 1 Tyr (Y)

three of the eight docking residues are preserved- Tyr7 and Tyr171 in the putative amino terminal binding pocket and Thr143 in the carboxy terminal binding pocket (Bahram et al, 1994). It is debatable whether the putative MIC molecules are capable of associating with peptide ligands, although sequence analysis indicates that the MIC molecules may fold in a similar manner to the class I chains despite being highly divergent.

3.3.6.4 Comparison Of The CD1 α3, Class I α3 and Class II β2 Domains.

The CD1 genes are members of the immunoglobulin supergene family according to various characteristics. For example, the derived CD1 amino acid sequences contain a pair of cysteine residues in both the α2 and α3 domains which can putatively form disulphide bridges in a manner analogous to the immunoglobulin domain loops. There is also a highly characteristic tryptophan residue (W) in the α3 domains at position 217 in a homologous position to that in the immunoglobulin chains (see Fig. 4).

The human CD1 molecules have been found to be approximately equally related to the class I and class II MHC molecules, suggesting that the CD1 genes arose early in evolution (~80mya) at a time close to the divergence of the class I and class II gene ancestors (Milstein et al, 1987; Martin et al, 1986; Calabi and Milstein, 1986). Comparison of the ovine CD1 α3 domains to the class I α3 and class II β2 domains (Table 7) produced similar results. The data indicate that if anything, the ovine CD1 α3 domains are slightly more related to the class II β2 domains than to the class I α3 domains (~30-33% identity with class II β2 compared to 25-30% identity with class I α3 domains). The α3 domain alignments in Figure 4 illustrate the similarities and differences between the ovine CD1 molecules and various other CD1, class I and class II sequences. The sheep data confirms the observations made in the human CD1 system by Milstein and colleagues (Milstein et al, 1987; Martin et al, 1986) that the CD1 molecules are approximately equally related to the class I and class II molecules. Detailed phylogenetic analysis would be required in order to comment in further detail on the CD1/MHC relationship.

3.3.6.5 Summary And Conclusions.

One of the main points which has emerged during this analysis is that the data support the

Table 7: Comparisons Between Homologous Domains Of Ovine CD1, MHC Class I and MHC Class II Amino Acid Sequences.

The results of pairwise amino acid sequence comparisons between the ovine CD1 $\alpha 3$ domains, ovine class I $\alpha 3$ domains (S class I) and ovine class II $\beta 2$ domains (S class II) are shown in the table. The genebank accession numbers for the sheep sequences used in this analysis are given in Appendix 3.

The figures in bold type are percentage identity results.

The figures in plain type are percentage similarity results.

	SCD1A25		SCD1B-42		SCD1T10	
S class I-1	**27.28**	51.11	**26.27**	46.67	**26.67**	47.78
S class I-2	**24.44**	48.49	**24.45**	45.56	**27.27**	51.14
S class I-3	**27.28**	51.11	**26.67**	47.78	**28.41**	53.41
S class I-4	**30.34**	53.94	**29.21**	50.56	**29.21**	50.56
S class II-1	**30.77**	50.55	**30.77**	51.65	**31.87**	49.45
S class II-2	**31.87**	54.95	**31.87**	53.85	**32.97**	56.05

view that SCD1A25, which contains an apparent single nucleotide deletion at position 602, encodes an expressed gene. The SCD1A25 amino acid sequence contains seven cysteine residues, six of which are conserved in position in all of the ovine CD1 molecules.Similarly, three of the four N-linked glycosylation sites in the classical CD1 sequences are conserved in all of the ovine sequences. A fourth site (258-260 in the α3 domain of HCD1B) which is specific to the HCD1B sequence is also conserved in SCD1A25. Furthermore, the SCD1A25 data obtained from the analysis of β_2m contact sites, CD8 interaction residues and class I peptide anchor residues did not differ from the results obtained with the other three ovine CD1 sequences. Taken together, these results imply that the single base deletion at position 602 in the SCD1A25 nucleotide sequence is an artefact which probably arose during the cDNA synthesis stage in preparation of the thymocyte library.

Previous studies have noted the divergence between CD1 molecules and MHC class I molecules despite their similar domain structure and association with CD1 (Calabi et al, 1989a; Balk et al, 1989). Recent developments in the analysis of MHC class I crystal structure have allowed a more thorough analysis of the interactions of ovine CD1 alpha chains with β_2m, CD8 and peptide ligands to be carried out. The results indicate that CD1 α chains interact with β_2m in a non-MHC class I-like manner. This may be interpreted to imply that the conservation of the α3 domains in the CD1 family is influenced by factor(s) other than the necessity for interaction with β_2m. An alternative view is that residues in the α3 domain of ovine CD1 molecules are conserved in order to preserve sites of interaction with β_2m, but that these sites differ from those utilised by MHC class I α chains. This latter view is supported by the biochemical evidence for interactions between CD1 α chains and β_2m (detailed in the literature review). The ovine CD1 sequences also lack typical CD8 binding residues and residues which are crucial for the docking of peptides in the groove. Taken together, the data suggest that the ovine CD1 molecules may not function in the same manner as MHC class I molecules to present peptide antigen to T cells, and indeed, there has only been one report to date which provides evidence for the involvement of exogenous antigen in CD1-specific T cell activation (Porcelli et al, 1992).

3.4 Southern Hybridisation Analysis of Ovine CD1.

3.4.1 Introduction.

The complexity of the CD1 gene family has been investigated in several species. Southern blot analysis of Molt4 DNA digested with EcoR I and probed with the insert from clone FCB6 (Calabi and Milstein, 1986) identified five hybridising bands (Martin et al, 1986). These bands were found to correspond by size to the EcoR I fragments of five different clones isolated from a genomic DNA library and represent the five different human genes, HCD1A, B, C, D and E. In the mouse, two highly homologous genes have been identified. The first gene, MCD1.1, was isolated by screening a genomic library with the insert from clone FCB6 (Bradbury et al, 1988). The second gene, termed MCD1.2, was isolated by screening a genomic library with a subclone of MCD1.1 which spanned the $\alpha3$ domain. In the domestic rabbit, the complexity of genomic Southern blots provides evidence for the existence of eight or more CD1 genes (Calabi et al, 1989b). CD1 genes have also been detected by Southern blot analysis in other mammalian species including cattle, rats and monkeys (Calabi et al, 1991).

In order to determine the complexity of the ovine CD1 gene family, a Southern blot analysis was performed using a homologous $\alpha3$ domain probe.

3.4.2 Southern Blot Analysis.

Genomic DNA was isolated and Southern blot analysis performed as described in section 2.11. A homologous $\alpha3$ domain probe was used to detect the CD1 genes as this domain displays the highest sequence homology among the CD1 genes of various species (76-87% nucleotide identity among human, mouse, rabbit and sheep genes). The homologous probe was derived from the SCD1T10 clone by PCR amplification (see section 2.14), as outlined in Figure 1B and C. The various restriction endonucleases used to digest the genomic DNA were selected such that none of the digestion sites were located within the $\alpha3$ domain of the known ovine sequences (Fig.1A). This was accomplished by analysing the restriction endonuclease digestion sites of the ovine $\alpha3$ domain sequences using the MAP programme

Figure 1:

A. Diagrammatic Explanation of Southern Blot Protocol.

The top line in the diagram is a linear representation of a CD1 molecule with the domain boundaries indicated by vertical dashed lines. Below this are linear representations of the three ovine CD1sequences which encode $\alpha 3$ domains- SCD1A25, SCD1B-41 and SCD1T10. The SCD1B-52 sequence has been omitted as it does not encode an $\alpha 3$ domain. The arrowheads indicate the positions of various restriction enzyme sites, numbered as follows:

1. Acc I	7. Hinc II
2. Xba I	8. Bam HI
3. Eco RI	9. Eco RV
4. Pst I	10. Sac I
5. Stu I	11. Bgl II
6. Ssp I	12. Bgl I

Note that the Stu I site (5) in SCD1A25 is only 11 nucleotides downstream from the 5' end of the $\alpha 3$ domain. Similarly, the Eco RV site (9) in SCD1T10 is only 17 nucleotides upstream from the 3' end of the $\alpha 3$ domain.

B and C. Preparation of The PCR Derived Homologous $\alpha 3$ Domain Probe.

Panel B shows a partial sequence of the insert from the SCD1T10 cDNA clone, with the borders of the $\alpha 3$ domain indicated by the arrowheads.

The primers G5936 (on the sense strand) and G5935 (on the antisense strand), as indicated in bold type, were used to amplify the $\alpha 3$ region using the PCR technique.

C. The photograph shows 10% of the PCR product analysed by gel electrophoresis. Molecular weight markers are indicated in base pairs.

A

5'UT | L | ALPHA 1 | ALPHA 2 | ALPHA 3 | TM/CYT | 3'UT

SCD1A25 — 1 2 3, 4, 5, 6

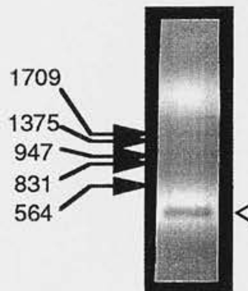
SCD1B-42 — 7, 1, 8, 9, 9 4, 8

SCD1T10 — 4 3, 9, 9, 7, 10, 11

B

```
                                  G5936
301  GCAAGTGAAG CC TGAAGCCT GGCTGTCCAG T GGCCCCACT CCTGGGCCTG
351  GCCGCCTATT GCTGGTCTGC CATGTCTCAG GATTCTATCC AAAACCTGTA
401  CGGGTGATGT GGATGAGGGG TGAGCAGGAG CAGCCTGGTA CTCAGCAAGG
451  AAACATCATA CTCAATGCTG ATTGGACTTG GTATCTCCGA GTAACCCTGG
501  ATGTGGCGGC TGGGGAGGCA GCTGGCCTGA GTTGTCGAGT GAAGCACAGC
551  AGTCTAGGAG ACCA GGATAT CATCCTGTAC TGGG GACACC CCATGTACAT
                               G5935
```

C



of UWGCG version 7.3 (Devereux et al, 1984). The two enzymes shown in Fig. 1A which cut just within the $\alpha 3$ domain of two of the ovine sequences were not considered problematic, since the overlaps with the probe sequence were minimal.

3.4.3 Results and Discussion.

Figure 2A shows the results of the overnight digestion of the genomic DNA analysed by gel electrophoresis. The observation of distinct bands in many of the tracks reflects the existence of satellite sequences in the DNA and provides evidence that the DNA had been thoroughly digested. Figure 2B shows a Southern blot of these genomic digests probed with the homologous $\alpha 3$ domain probe. The results indicate the existence of up to seven ovine CD1 genes, assuming at least one cut site per enzyme between each gene. The enzymes used were carefully selected to minimise the risk of observing more than one band derived from a single gene. However, the number of CD1 genes cannot be stated precisely, due to the possible existence of enzyme cut sites within the $\alpha 3$ domains of genes which have not yet been isolated and sequenced.

Figure 2: Analysis of Ovine CD1 by Southern Hybridisation of Genomic DNA.

Panel A shows the results of overnight digestion of sheep genomic DNA with a variety of restriction endonuclease enzymes, with lanes numbered as follows:

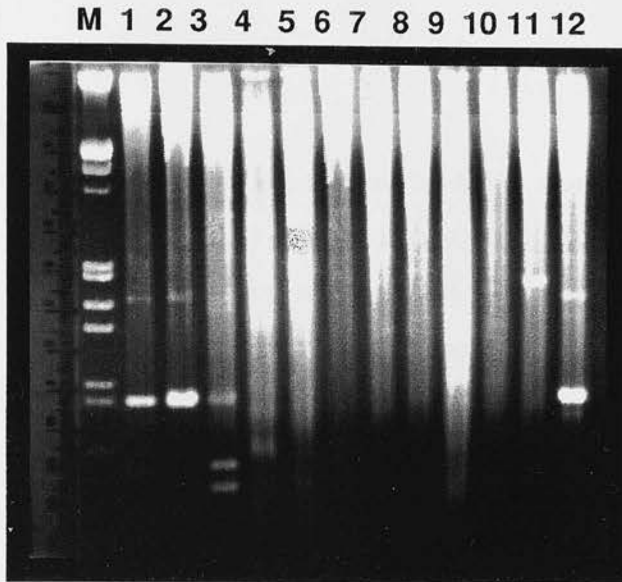
1. Sac I	7. Bgl I
2. Bam HI	8. Xba I
3. Eco RI	9. Ssp I
4. Pst I	10. Acc I
5. Hinc II	11. Stu I
6. Eco RV	12. Bgl I

Approximately 10μg of genomic DNA was digested with each enzyme and analysed by gel electrophoresis. Molecular weight markers (Eco RI/Hind III cut λ DNA) are indicated in the left hand lane.

Panel B shows the result of Southern hybridisation of the DNA digests shown in panel A. The DNA was blotted onto nylon membrane and probed with the homologous α3 domain probe.

Molecular weight markers are indicated in kilobase pairs.

A



B

21.23
5.15
4.97
4.27
3.53
2.03
1.90
1.58
1.38
0.95
0.83
0.56

3.5 PCR Analysis Of CD1 mRNA Transcription In The Intestinal Mucosa.

3.5.1 Introduction.

All four ovine cDNA clones were isolated from thymocyte cDNA libraries and were found to be homologues of the human CD1B gene. Various studies have additionally detected CD1 expression, at both the mRNA level and on the cell surface, in tissues other than the thymus in both humans and mice, and in particular, CD1 expression has been observed in the intestinal mucosa.

In the mouse, the highest levels of transcription of the two CD1D-type genes were detected in the thymus, liver and spleen, and much lower, probably insignificant, levels were detected in the kidney and brain (Bradbury et al, 1988). MCD1.2 transcripts were approximately five-fold less abundant than those of MCD1.1 in all tissues except for the thymus where mRNA levels were equal. In a similar study, murine CD1 message was low or undetectable in the thymus by Northern blot analysis (Balk et al, 1991a). Bleicher and colleagues (1990) detected murine CD1 expression at the cell surface in the thymus and intestine, and cytoplasmic expression in liver hepatocytes, using the anti-murine CD1.1 monoclonal antibodies 1H1 and 3C11 to stain tissue sections. Similarly, Mosser and coworkers (1991) used a rabbit antimurine CD1.2 polyclonal antiserum to analyse CD1 expression in mouse tissues. They found that CD1 was expressed in the thymus, spleen, liver and lungs of adult mice and additionally in the brain, kidney, heart and intestine of foetal mice. The discrepancies in detection of intestinal CD1 expression have been resolved by *in situ* hybridisation analysis, which indicated that only Paneth cells in the epithelium express CD1 message at high levels (Lacasse and Martin, 1992).

In humans, the protein products of the CD1A, B and C genes are differentially expressed on immature thymocytes, Langerhans cells, dendritic cells and a subset of B cells (Calabi et al, 1991). In addition, expression of the CD1D gene product has been detected predominantly on intestinal epithelial cells. Human CD1d expression was detected in the intestinal

epithelium using the anti-murine CD1.1 monoclonal antibodies 1H1 and 3C11 as well as an HCD1d-specific antipeptide antiserum. A human intestinal $CD8^+$, $\alpha\beta^+$ T cell line was found to exhibit significant CD1-specific cytolytic activity (Balk et al, 1991). The monoclonal antibody 3C11 was additionally found to inhibit $CD8^+$T cell proliferation in a mixed lymphocyte reaction (MLR) where the stimulator cells were normal intestinal epithelial cells but did not cause inhibition in a conventional MLR (Panja et al, 1993). Most recently, human CD1d was found to be expressed in the intestinal epithelium as a 37kD protein that was β_2m-independent with no N-linked carbohydrate (Balk et al, 1994).

Overall, the studies in both humans and mice indicate that CD1 expression is not confined to the thymus, and that CD1d expression is predominantly detected in the intestine. Since the ovine cDNA sequences were all isolated from thymocyte libraries, it was considered important to investigate CD1 expression in an extra-thymic site. Consequently, the PCR technique has been employed in an attempt to detect ovine CD1 transcription in the intestinal epithelium as outlined below.

3.5.2 Experimental Rationale

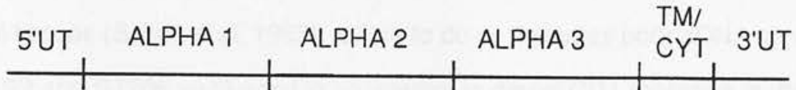
The simplest and quickest approach to the analysis of CD1 expression in the intestinal mucosa involved use of the PCR technique to detect CD1 gene transcription. The primers used were the degenerate 5' primer B199 (see section 3.2, Fig. 2B) in conjunction with the 3' primer G7206. Primer G7206 is based on a conserved region located fourteen nucleotides downstream of the translational stop codon in all four ovine cDNA clones (Fig. 1). In theory, this primer combination would amplify any CD1 specific cDNA present in an intestinal sample, assuming the conservation of both primer sequences in all ovine CD1 mRNA transcripts. Subsequent blotting and hybridisation of any PCR product(s) with an $\alpha3$ domain probe would confirm that the product was CD1. In addition, hybridisation with clone-specific probes would indicate whether the PCR product represented a previously isolated ovine CD1 or an alternative CD1 sequence. Any PCR products could be cloned and sequenced to confirm the identity of the amplified product.

Figure 1: Derivation And Use Of The Primer G7206 in Analysis Of CD1 Transcription In The Intestinal Mucosa.

A. The diagram shows a linear representation of a CD1 molecule with the domain boundaries indicated by vertical lines. The approximate locations of the 5' degenerate primer B199 and the 3' primer G7206 are also indicated.

B. The aligned sense strand sequences illustrate the derivation of the 3' antisense primer G7206. The highly conserved 18mer sequence is located fourteen nucleotides downstream of the translational stop codon in the 3' untranslated region. The nucleotide numbering refers to the SCD1B-42 sequence in Fig. 1, section 3.3.

A



B199 G7206

B

	1114 ... 1130
SCD1B-52	GTCTCCTTTTCCATTTGG
SCD1B-42	GTCTCCTTTTC.ATTTGG
SCD1T10	GTCTCCTTTCCCATTTGG
SCD1A25	GTCTCCTTTTCCATTTGG

	5' ... 3'
G7206	CCAAATGGAAAAGGAGAC

3.5.3 PCR Protocol For Analysis of Intestinal CD1 Transcription.

Total RNA was isolated from sheep ileum and thymus as described in section 2.12. cDNA was synthesised using random hexanucleotide primers in order to overcome potential annealing problems with the use of specific primers, as discussed previously (see section 3.2.2 and 3.2.5). The suitability of the cDNA for use in subsequent PCR reactions was tested by performing an ATPase PCR reaction on all samples. The ATPase message should be detectable in all tissues (Schull et al, 1985)- failure to do so indicates poor cDNA synthesis. The primers B199 and G7206 were used in an attempt to detect CD1 message in the intestinal mucosa. Two rounds of PCR amplification were performed- the template for the second round of PCR was 3μl of product from the first round and the same primers were used in the second round reaction. Details of all PCR reactions are given in section 2.14. The sequences of the ATPase, B199 and G7206 primers used are listed in Appendix 2B. Control PCR reactions included negative controls where the cDNA template was replaced with an equal volume of SDW, and positive controls using a thymus cDNA template and a sample of double-stranded plasmid DNA from the SCD1B-52 clone (the SCD1B-52 clone insert is the only ovine clone which contains both the 5' and 3' primer sites).

A single band of the expected size (approximately 1000bp) was detected in the ileum sample following two rounds (70 cycles) of PCR amplification (Fig. 2A, lane 3). Control PCR reactions gave the expected results with the exception of the thymus tissue positive control sample which was negative (Fig. 2A, lane 2). However, the second positive control which employed double-stranded SCD1B-52 plasmid DNA as the template produced a single band of the expected size (~850bp, Fig. 2A, lane 1), suggesting that the primers were amplifying the appropriate target sequence. It seems that there may have been a problem with the cDNA synthesis step, although the ATPase positive controls gave the expected results. The single band produced in the ileum sample was cloned using the TA Cloning kit (Invitrogen). Several insert-positive clones were obtained for sequencing. The results of Eco RI digests of four plasmid DNA samples are shown in Fig. 2B. Two insert bands were observed in each case (~650bp and 350bp), indicating the presence of an additional Eco RI site within the insert. The insert from a representative clone was sequenced on each strand with the M13 forward

Figure 2: PCR Analysis of CD1 Transcription In The Intestinal Mucosa.

A. The photograph shows the results of PCR amplification using the primers B199 and G7206 as outlined in the text. 10% of the PCR products from the second round of PCR were analysed by gel electrophoresis.

Lane 1: Positive control PCR using SCD1B-52 dsDNA as the template.

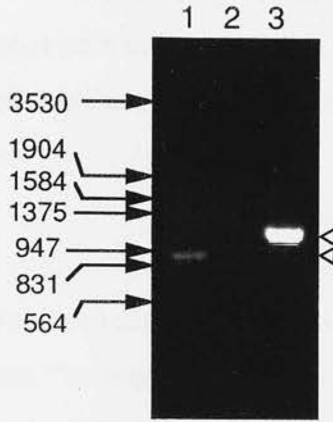
Lane 2: Positive control PCR using thymus cDNA as the template.

Lane 3: PCR using ileum cDNA as the template.

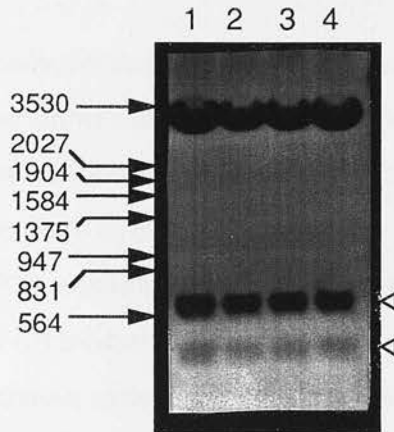
B. Eco RI digests of dsDNA from four plasmid samples analysed by gel electrophoresis after cloning the ileum PCR product using the TA cloning kit.

Molecular weight markers are indicated in base pairs.

A



B



and reverse sequencing primers (Appendix 2A) using the dideoxy chain termination technique (Sanger et al, 1977). The initial sequence data for each strand revealed a G7206 primer site at each end of the insert. Furthermore, the initial sequences showed no homology to known CD1 sequences, when compared using the GAP programme of UWGCG 7.3. A search of the genEMBL database using the FASTA programme of UWGCG 7.3 revealed that the ileum PCR sequence gave the highest score when compared to the *C. elegans* N-acetylgalactosyaminyltransferase sequence. No further sequence analysis was performed for the ileum PCR samples.

3.5.4 Discussion.

The PCR technique has been used in an attempt to detect transcription of ovine CD1 genes in the intestinal mucosa. The single band obtained by PCR amplification from an ileum cDNA sample was cloned and partially sequenced. Pairwise comparisons of the initial sequence data for the cloned PCR product against the human and ovine CD1 sequences gave percentage identity values in the range 37-45%. The results indicate that the PCR product represents an inadvertent amplification of a sequence other than the target sequence, as confirmed by a FASTA search of the genEMBL database.

It is not possible to comment on CD1 expression in the intestinal mucosa on the basis of the above results, and no further experiments were performed given the limited time available. However, various alternative strategies could be pursued in order to investigate this further. Specific PCR primers could be derived for each ovine CD1 clone and the specificity confirmed by positive control PCR reactions using clone DNA and thymus cDNA. Blotting and hybridisation of the PCR products with an $\alpha 3$ domain probe would confirm that they were CD1 sequences. These primers could subsequently be used to analyse CD1 expression in the intestinal mucosa by PCR. A negative result would then be interpreted as lack of expression of CD1 mRNA corresponding to the ovine CD1 sequences already isolated. However, this procedure would not allow the cloning and sequencing of a previously unknown ovine CD1 sequence, in contrast to the PCR protocol outlined above (3.5.2). The results of this procedure may be complicated by the potential specificity of the PCR primers for any

unknown transribed CD1 genes.

An alternative method for investigation of CD1 gene transcription would be Northern blot analysis using an $\alpha 3$ domain probe (which should hybridise to all transcribed CD1 genes) in conjunction with clone specific probes (to assign transcripts to the known ovine CD1 genes). This analysis would enable the number, variability and relative abundance of CD1 transcripts in the intestine and other tissues to be determined.

A third possible approach to an analysis of CD1 mRNA transcription would be *in situ* hybridisation (ISH) using an $\alpha 3$ domain probe and clone specific probes as mentioned above. This technique allows the detection of gene products which are not necessarily expressed at a high level on the cell surface and which may only be expressed on a subset of cells within a complex tissue.

Recent studies in the sheep using a panel of mAbs raised against sheep and cattle thymocytes, DC's and intestinal epithelial cells have indicated that ovine CD1 is, in fact, expressed in the intestinal mucosa, although it is not yet clear which cell types within the tissue are positive for CD1 staining (S. Rhind, personal communication).

4. Discussion.

The CD1 molecules are a family of β2m-associated glycoproteins with broad structural homology and weaker sequence homology to the MHC class I antigens. The CD1 molecules are characterised by limited polymorphism, low surface expression and a restricted pattern of tissue distribution in comparison to the classical MHC class I gene products.

CD1 genes have now been characterised in humans (Calabi et al, 1991), mice (Bradbury et al, 1988), rabbits (Calabi et al, 1989b) and most recently in the rat (Ichimiya et al, 1994). CD1 molecules have additionally been identified by monoclonal antibodies in cattle (MacHugh et al, 1984; Howard et al, 1993 a and b), pigs (Pescovitz et al, 1984; Pescovitz et al, 1990) and in the sheep (Mackay et al, 1985; Dutia and Hopkins, 1991; Bujdoso et al, 1989).

4.1 The Ovine CD1 Gene Family.

To date, the limited information available for the ovine CD1 family has been obtained using a small panel of anti-bovine and anti-ovine CD1 mAbs in a series of biochemical and histological assays. All mAbs were found to detect a 46kD α chain in association with β2m. In order to extend our understanding of these molecules, the ovine CD1 family has now been characterised at the molecular level.

Four ovine CD1 clones have been isolated by library screening of thymocyte cDNA libraries. The first clone identified, SCD1A25, was isolated using a human CD1C α3 region probe. A homologous probe obtained by PCR amplification of the α3/TM/CYT region of SCD1A25 was subsequently utilised to isolate an additional three ovine clones, SCD1B-42, SCD1B-52 and SCD1T10. Interestingly, all four ovine clones are most homologous to the human CD1B gene. Southern blot analysis has indicated the existence of up to seven ovine CD1 genes and it is possible that several expressed ovine CD1 genes remain to be identified. Three of the four clones (SCD1A25, SCD1B-42 and SCD1T10) contain inserts truncated at the 5' end. The fourth clone (SCD1B-52) contains an insert which is intact at the 5' end but has a precise in-frame deletion of the α3 domain. The missing 5' sequences (part of the 5' untranslated region as well as the α1 domain) for the SCD1A25 and SCD1B-42 clone inserts were

obtained by PCR amplification.

The ovine CD1 sequences were compared to the human, murine and rabbit sequences. In addition, a detailed structural analysis of the interactions of ovine CD1 α chains with β_2-microglobulin, CD8 and peptide was performed based on the data obtained from crystal structure analysis of the MHC class I molecules. The results indicate that the CD1 α chains apparently interact with β_2-microglobulin in a non- MHC class I like manner. The ovine CD1 sequences also lack typical CD8 binding sites and residues which are crucial for the docking of peptides in the MHC class I groove. Taken together, the data imply that CD1 molecules may not function in an identical manner to the class I molecules to present peptide antigen to T cells.

Several areas of future work have been suggested by data reported in this thesis. It will be necessary to pursue appropriate experiments in order to correlate the results from the mAb and molecular work. This could be achieved by transfecting appropriate cells with constructs containing the cDNA for both SCD1B-42 and SCD1B-52 and screening transfected cells by FACS analysis with the available mAbs. In the case of SCD1B-42, it would first be necessary to ligate the 5' UT/α1 DNA sequence obtained by PCR amplification to the main cloned cDNA sequence. In order to analyse SCD1T10 in this system, it would first be necessary to obtain DNA sequence for the 5'UT and α1 domains. It would only be possible to analyse SCD1A25 using this procedure by first of all rescreening a thymocyte cDNA library in order to obtain a true, in-frame cDNA clone, since the currently available cDNA clone contains a single nucleotide deletion. Analysis of SCD1B-52 in this system could prove interesting, since SCD1B-52 does not contain the α3 domain. This is the main site of interaction with β_2m, and it is now known that the human CD1d molecule does not require interaction with β_2m for cell surface expression (Balk et al, 1994). It may also be possible to confirm SCD1B-52 membrane expression by generating an antipeptide antiserum against an epitope encoded by the junction of the α2 and TM domains, although a negative result obtained when screening with such an antiserum would not necessarily imply a lack of SCD1B-52 cell surface expression.

The characterisation of the ovine CD1 molecules at the molecular level will now permit the generation of molecular probes for use in *in situ* hybridisation and Northern blot analysis of CD1 transcription in various tissues. Use of an α3 domain probe in conjunction with clone specific probes (for example, an α2/TM probe specific for SCD1B-52) in a Northern blot would enable the number, variability and relative abundance of CD1 transcripts in various tissues to be determined. These probes could also be employed in *in situ* hybridisation which would enable a detailed analysis of CD1 cell and tissue expression to be performed. Such experiments would be particularly interesting in intestinal tissue, since CD1d is predominantly expressed on human intestinal epithelium and has been implicated in antigen presentation to $CD8^+$ T cells at this site (Balk et al, 1991; Panja et al, 1993).

4.2 Evolutionary Relationship Between CD1 and MHC Molecules.

Two major theories have been proposed to explain MHC evolution. One theory holds that class I molecules evolved first as the result of a recombination event between an Ig-like C domain and the peptide binding domain of an HSP70 molecule (Flajnik et al, 1991, 1993). This model was proposed as a consequence of noting the similarities between class I α peptide binding domains and the peptide binding region of members of the HSP70 family (Ripmann et al, 1991). The second major theory of MHC evolution holds that class II molecules evolved first, and that class I molecules subsequently evolved as a result of a recombinatorial event. These two theories are discussed in further detail below and form the context for a discussion of CD1 evolution.

The class I first theory of MHC evolution developed by Flajnik and colleagues was primarily based on the observations and ideas of the Milstein group. Calabi and Milstein (1986) first proposed that the CD1 genes arose before the separation of the primate and rodent branches around the same time as the divergence of the class I and class II ancestor genes when they analysed the CD1A gene. They noted that although both the CD1 and class II β genes had diverged from class I α genes to an almost equal extent, only the CD1 genes retain the interaction with β_2-microglobulin. This theory was further expanded by the Milstein group (Martin et al, 1986; Milstein et al, 1987) when they suggested that the large divergence

between the CD1 and MHC families must have involved a duplication of an ancestral gene and subsequent transposition to a different chromosome around the time of the class I/class II separation. They proposed that the ancestral precursor of all MHC-type genes was probably β_2-microglobulin associated i.e. class I-like, and that the CD1 genes may have arisen from the class II ancestor after it separated from class I but before it lost its association with β_2m.

Flajnik and colleagues proposed a model for MHC evolution basing their ideas on those already discussed by Martin, Milstein et al. Essentially, this group had cloned a Xenopus class I MHC molecule (Flajnik et al, 1990) which, on sequencing, was found to have some similarity to the human HSC70 sequence (24% identity with gaps inserted). HSC70 is a member of the HSP70 heat shock protein family, molecules which stimulate a high proportion of $\gamma\delta$ T cells (Porcelli et al, 1991; Born et al, 1990) and which have been implicated in presentation of antigenic peptides to T cells, possibly through "feeding" of the peptides to MHC molecules (Vanbuskirk et al, 1989). Despite the low homology between the Xenopus class I sequence and HSC70, further analysis proved interesting. Secondary structure predictions indicated that the Xenopus class I and HSC70 binding domain sequences conform to the structure of human class I predicted by crystal structure analysis. Hydropathy plots suggested that many hydrophobic and hydrophilic amino acids are similar between the class I and HSC70 sequences. In addition, several features of HSC70 were mapped to homologous positions in the class I crystal structure. The authors concluded that the peptide binding domains of class I and HSC70 may adopt a similar symmetrical structure and on this basis proposed that genes encoding a peptide binding structure were transferred en masse to a gene segment encoding an immunoglobulin domain during MHC evolution. This scenario is illustrated in Figure 1, which shows the proposed evolution of the CD1 genes from the class II gene ancestor.

The second major hypothesis on MHC evolution holds that the class II MHC evolved first and that the class I α chain arose from a recombinatorial event that brought together in one gene the exon encoding a class II α1 domain and the exons encoding the class II β1 and β2 domains (Lawlor et al, 1990). This model was originally proposed on a thermodynamic basis in

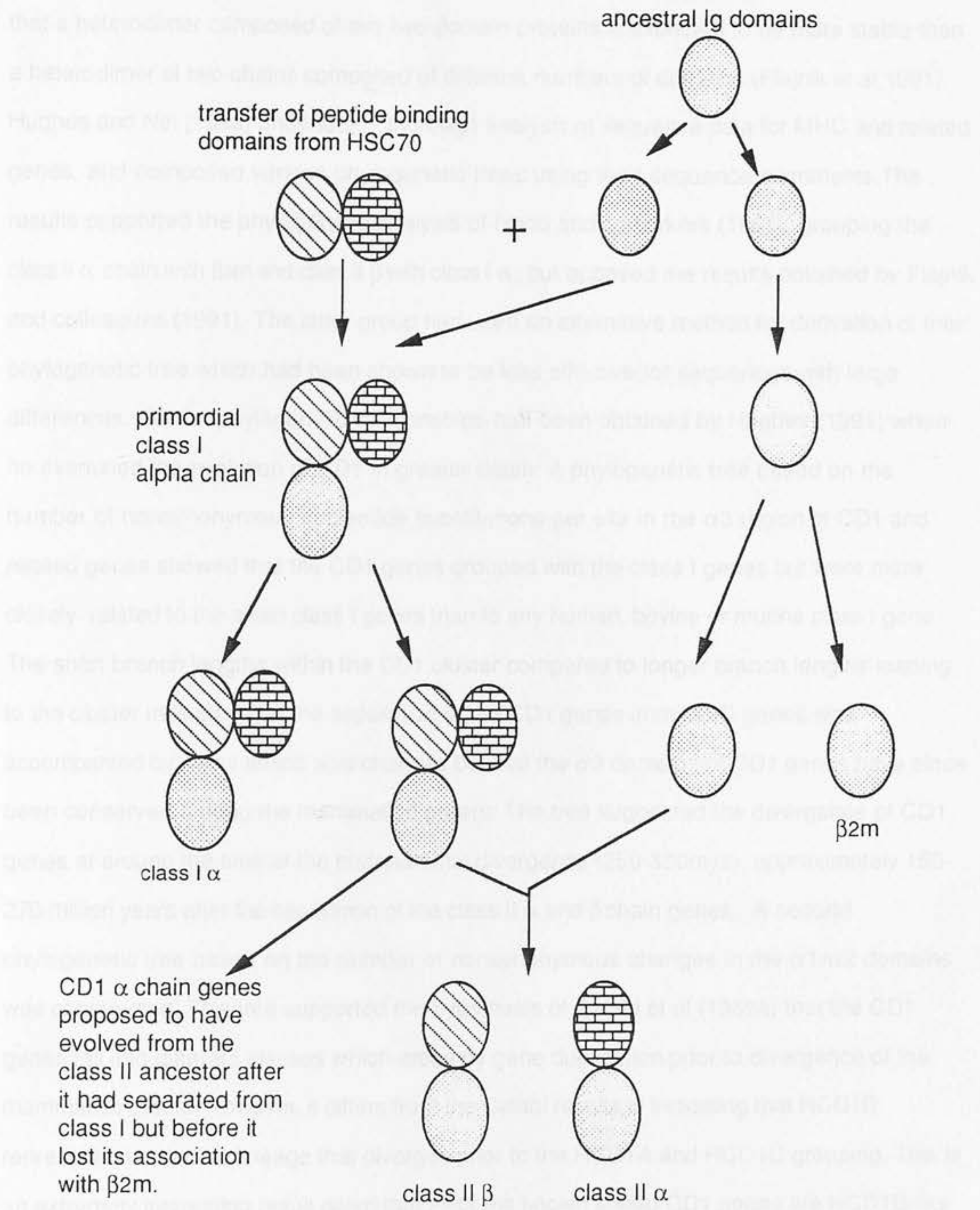
Figure 1: Evolution Of MHC and CD1 Molecules From An MHC Class I-like Ancestor.

The diagram illustrates schematically the model of MHC and CD1 evolution proposed by Flajnik and coworkers and the Milstein group (see text). This model holds that MHC class I molecules evolved first as a result of a recombination between an immunoglobulin-like C domain and the peptide binding domains of an HSP70 heat shock protein. The class I gene and the gene ancestral to β_2m then duplicated in tandem, followed by a recombination event which brought together the $\alpha 1$ domain of the three-domain gene with the C domain gene thus giving rise to the class II α chain gene. The three-domain gene which lost the $\alpha 1$ domain became the two-domain ancestor of class II β. In this model, it was proposed that the CD1 genes evolved from the class II ancestor after it had separated from class I but before it lost its association with β_2m.

The dotted domains represent domains homologous to class I $\alpha 1$.

The striped domains represent domains homologous to class I $\alpha 2$.

The blocked domains represent domains homologous to class I $\alpha 3$ (i.e. to immunoglobulin C domains).


ancestral Ig domains
transfer of peptide binding
domains from HSC70
+
primordial
class I
alpha chain
class I α
class II β
class II α
β2m
CD1 α chain genes
proposed to have
evolved from the
class II ancestor after
it had separated from
class I but before it
lost its association
with β2m.

that a heterodimer composed of two two-domain proteins is expected to be more stable than a heterodimer of two chains composed of different numbers of domains (Flajnik et al,1991). Hughes and Nei (1993) undertook a thorough analysis of sequence data for MHC and related genes, and composed various phylogenetic trees using their sequence alignments.The results supported the phylogenetic analysis of Hood and coworkers (1985), grouping the class II α chain with β_2m and class II β with class I α, but opposed the results obtained by Flajnik and colleagues (1991). The latter group had used an alternative method for derivation of their phylogenetic tree which had been shown to be less effective for sequences with large differences. Similar phylogenetic relationships had been obtained by Hughes (1991) when he examined the evolution of CD1 in greater depth. A phylogenetic tree based on the number of nonsynonymous nucleotide substitutions per site in the α3 region of CD1 and related genes showed that the CD1 genes grouped with the class I genes but were more closely related to the avian class I genes than to any human, bovine or murine class I gene. The short branch lengths within the CD1 cluster compared to longer branch lengths leading to the cluster indicated that the separation of the CD1 genes from MHC genes was accompanied by many amino acid changes but that the α3 domains of CD1 genes have since been conserved among the mammalian orders. The tree suggested the divergence of CD1 genes at around the time of the bird/mammal divergence (250-300mya), approximately 150-270 million years after the separation of the class II α and β chain genes. A second phylogenetic tree based on the number of nonsynonymous changes in the $\alpha1/\alpha2$ domains was constructed. This tree supported the hypothesis of Calabi et al (1989a) that the CD1 genes fall into different classes which arose by gene duplication prior to divergence of the mammalian orders. However, it differs from the Calabi results in indicating that HCD1B represents a separate lineage that diverged prior to the HCD1A and HCD1C grouping. This is an extremely interesting result given that all of the known sheep CD1 genes are HCD1B-like. The Hughes model of evolution is summarised in Figure 2.

The majority of the evidence thus supports the hypothesis that the ancestral MHC molecule had a class II-like structure and that the CD1 molecules diverged ~150-270 million years after the separation of the class II α and β chain genes.

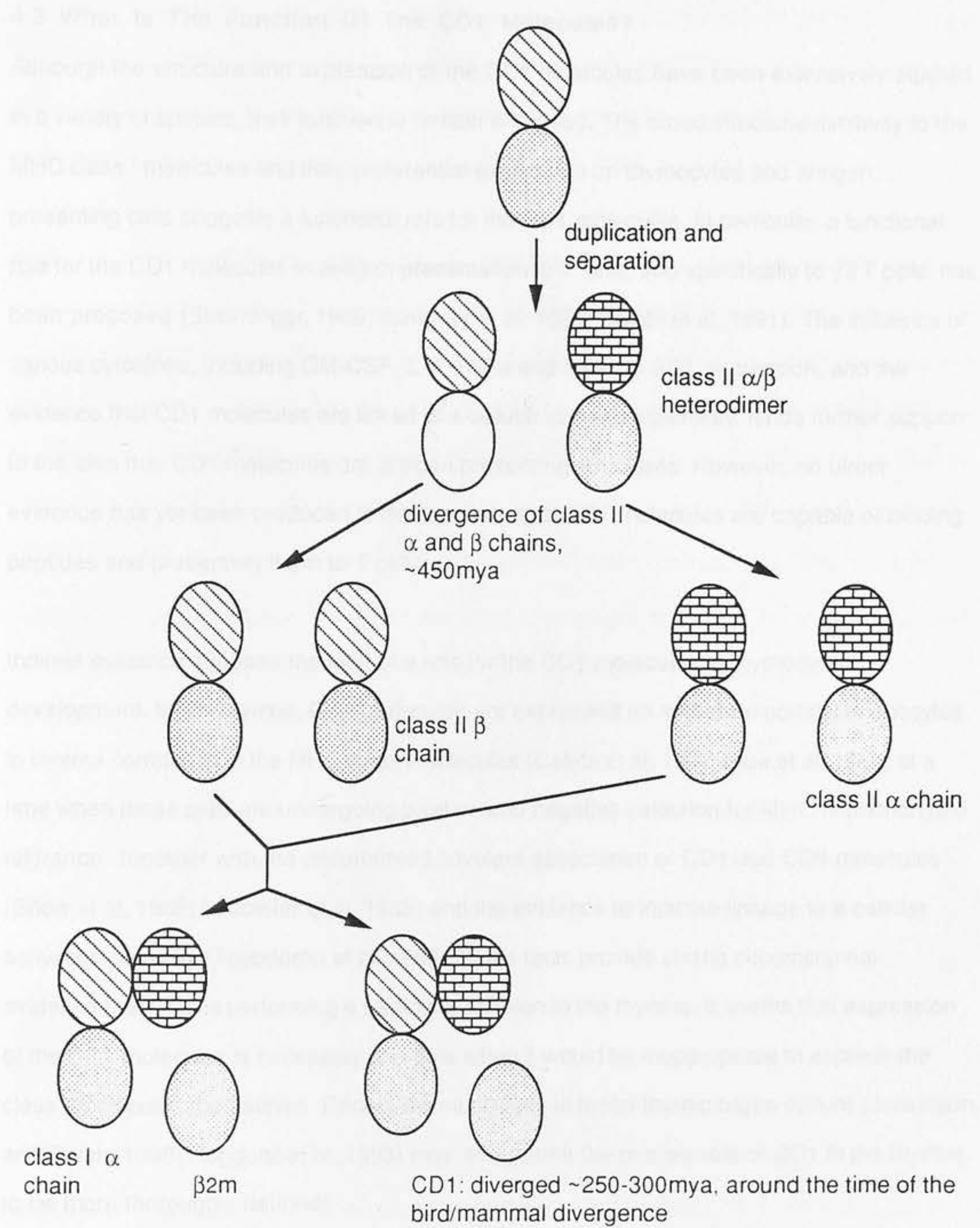
Figure 2: Evolution Of MHC and CD1 Molecules From An MHC class II-like Ancestor.

The diagram illustrates schematically the model of MHC and CD1 evolution proposed by Hughes and colleagues. This model holds that MHC class II molecules evolved first. The two-domain ancestral class II gene duplicated giving rise to the class II α/β heterodimer. The α and β chain genes then duplicated in tandem. The class I α chain gene arose by the transfer of an $\alpha1$ domain exon to an α chain gene. the loss of this exon left a single domain gene which became the β_2m ancestor. In this model, the CD1 α chain genes are proposed to have arisen from the class I ancestor and diverged from the class I α chain genes at around the same time as the bird/mammal divergence (250-300Mya).

The dotted domains represent domains homologous to class $\alpha1$.

The striped domains represent domains homologous to class I $\alpha2$.

The blocked domains represent domains homologous to class $\alpha3$ (i.e. to immunoglobulin C domains).


duplication and separation
class II α/β heterodimer
divergence of class II α and β chains, ~450mya
class II β chain
class II α chain
class I α chain
β2m
CD1: diverged ~250-300mya, around the time of the bird/mammal divergence

4.3 What Is The Function Of The CD1 Molecules?

Although the structure and expression of the CD1 molecules have been extensively studied in a variety of species, their function(s) remain ill-defined. The broad structural similarity to the MHC class I molecules and their preferential expression on thymocytes and antigen presenting cells suggests a functional role for the CD1 molecules. In particular, a functional role for the CD1 molecules in antigen presentation to T cells, and specifically to $\gamma\delta$ T cells, has been proposed (Strominger, 1989; Janeway et al, 1988; Calabi et al, 1991). The influence of various cytokines, including GM-CSF, IL-4, TNFα and IL-6, on CD1 expression, and the evidence that CD1 molecules are linked to a cellular activation pathway, lends further support to the idea that CD1 molecules are antigen presenting structures. However, no direct evidence has yet been produced to demonstrate that CD1 molecules are capable of binding peptides and presenting them to T cells.

Indirect evidence supports the idea of a role for the CD1 molecules in thymocyte development. In the thymus, CD1 molecules are expressed on immature cortical thymocytes in inverse correlation to the MHC class I molecules (Calabi et al, 1991; Blue et al, 1989) at a time when these cells are undergoing positive and negative selection for MHC restriction and tolerance. Together with the documented covalent association of CD1 and CD8 molecules (Snow et al, 1985; Ledbetter et al, 1985) and the evidence to indicate linkage to a cellular activation pathway (Theodorou et al, 1990), these facts provide strong circumstantial evidence that CD1 is performing a particular function in the thymus. It seems that expression of the CD1 molecules is necessary at a time when it would be inappropriate to express the class I molecules themselves. Recent developments in foetal thymic organ culture (Jenkinson and Owen, 1990; Hogquist et al, 1993) may now permit the precise role of CD1 in the thymus to be more thoroughly defined.

The postulated role of CD1 in antigen presentation was originally proposed on the basis that the limited germline diversity of $\gamma\delta$ TCR's implied recognition of relatively nonpolymorphic ligands (Janeway et al, 1988; Strominger, 1989). However, it is now known that the $\gamma\delta$ TCR repertoire is actually larger than the $\alpha\beta$ TCR repertoire due to the great potential for junctional

diversity (Porcelli et al, 1991). The original simplified view of antigen presentation to γδ T cells has also been challenged by the discovery that nonpolymorphic ligands, such as the class Ib molecule Qa-2, are potentially capable of binding a diverse array of peptides and not just a limited set as was previously assumed (Rotzschke et al, 1993; Joyce et al, 1994).

Since αβ T cell recognition requires antigen presentation by MHC molecules, it has been commonly assumed that γδ T cell recognition also follows the same set of rules. However, this basic assumption has now been challenged by two recent reports which have suggested that the molecular nature of γδ T cell recognition is fundamentally different to that of αβ T cells (Schild et al, 1994; Sciammas et al, 1994). Schild and colleagues analysed the recognition of MHC molecules by the γδ T cell hybridomas LBK5, which is specific for the class II molecules $IE^{k,b,s}$, and G8 which recognises the nonclassical MHC class I molecule TL. The reactivity of neither clone required functional class I nor class II processing pathways in the stimulator cells. Furthermore, the LBK5 epitope was mapped distal to the peptide binding groove by analysing the response of the clone to a panel of thirteen antigen presenting cell lines and comparing the results to normal αβ T cell responses. Sciammas and colleagues (1994) obtained similar results when they investigated the stimulation requirements of the γδ T cell clone Tg14.4 which recognises a *Herpes simplex* glycoprotein in an MHC class I/class II independent manner. When expressed in the antigen processing mutant cell line RMA/S, the glycoprotein was capable of being recognised by Tg14.4. Furthermore, it was demonstrated that Tg14.4 can recognise whole unprocessed glycoprotein in the absence of any antigen presenting cells. Additional support for the view that γδ T cells do not recognise antigen in a manner analogous to αβ T cells was produced by Rock and colleagues (1994) when they characterised the length distribution of CDR3 regions in various immune receptor chains. The results indicated that the CDR3 lengths of both α and β TCR polypeptides are nearly identical and are highly restricted in length. This size constraint was attributed to the requirement for recognition of small peptides in the groove, since the CDR3 region is critical in antigen recognition (Davis and Bjorkman, 1988). In contrast, CDR3 lengths of immunoglobulin H chains are long and variable, while L chain CDR3's are shorter and less variable. For the γδ TCR, δ chain CDR3 lengths are long and variable, while those of γ chains are longer and more

constrained. In this respect, $\gamma\delta$ TCR's resemble the immunoglobulins in containing CDR3 regions of discordant length, whereas $\alpha\beta$ TCR's contain CDR3 regions which are virtually identical in length.

The recent developments in our understanding of antigen recognition by $\gamma\delta$ T cells have further clouded the original picture that the nonpolymorphic CD1 molecules may participate in peptide presentation to these cells. In the past decade, experiments designed to characterise the specificity and function of $\gamma\delta$ T cells have been based on our detailed knowledge of $\alpha\beta$ T cell biology. I view this as a "square peg in a round hole" scenario, given that almost a decade of intense investigation has failed to produce a consensus view on the specificity and function of $\gamma\delta$ T cells. Similarly, studies designed to elucidate the functions of the CD1 molecules can be viewed in this way i.e. it has been assumed that CD1 molecules, by virtue of their broad structural similarity to the class I molecules, will perform similar functions to the class I molecules. However, the discovery that the p51 molecule, an Fc receptor on intestinal epithelial cells of the neonatal rat, is a β_2m-associated class I-like structure has already demonstrated that class I-like structure does not necessarily imply a class I-like function (Simister and Mostov, 1989). A non-class I-like function for the CD1 molecules is further supported by the results from the detailed structural analysis of the ovine CD1 molecules. The analysis indicated that CD1 molecules do not possess the structural requirements for presentation of peptides in an MHC class I-like manner, although it cannot be concluded that they do not present peptides in some other manner.

In order to further elucidate the function(s) of CD1, it will be necessary to pursue several lines of investigation. In particular, recent developments in the ability to identify peptides eluted from MHC molecules should allow the question of the peptide binding capacity of the CD1 molecules to be resolved. Determination of the crystal structure of CD1 will permit the detailed structure of these molecules to be established. Development of transgenic mice which do not express CD1 may also help establish the precise role of CD1, particularly in the thymus and in intestinal tissue. Finally, it may be fruitful to investigate further the association of the CD1 molecules with intracellular cell signalling pathways, given the recent advances which have

been made in this field. In particular, there is increasing evidence to show that, in addition to their role in antigen presentation, MHC class II molecules can transduce signals which are important in regulating APC function during cognate T-B cell interactions (Scholl and Geha, 1994). It will therefore be interesting to determine whether class I molecules and other molecules implicated in antigen presentation, such as CD1 and class Ib molecules, can also deliver signals which modulate cell function.

References

Albertson, D.G., Fishpool, R., Sherrington, P., Nacheva, E. and Milstein, C. (1988). Sensitive and high resolution *in situ* hybridisation to human chromosomes using biotin labelled probes: assignment of the human thymocyte CD1 antigen genes to chromosome 1. *EMBO J.* **7**, 2801-2805.

Aldrich, C.A., Waltrip, R., Hermel, E., Attaya, M., Fischer Lindahl, K., Monaco, J.J. and Forman, J. (1992). T cell recognition of Qa-1^b antigens on cells lacking a functional Tap-2 transporter. *J. Immunol.* **149**, 3773-3777.

Allen, P. M. (1994). Peptides in positive and negative selection: a delicate balance. *Cell* **76**, 593-596.

Allison, J.P. and Lanier, L.L. (1987). Structure, function and serology of the T-cell antigen receptor complex. *Ann. Rev. Immunol.* **5**,503-540.

Allison, J.P. and Havran, W.L. (1991). The immunobiology of T cells with invariant $\gamma\delta$ antigen receptors. *Ann. Rev. Immunol.* **9**, 679-705.

Amigorena, S., Drake, J. R., Webster, P. and Mellman, I. (1994). Transient accumulation of new class II MHC molecules in a novel endocytic compartment in B lymphocytes. *Nature* **369**, 113-120.

Amiot, M., Bernard, A., Raynal, B., Knapp, W., Deschilde, C. and Boumsell, L. (1986). Heterogeneity of the first cluster of differentiation: characterisation and epitopic mapping of three CD1 molecules on normal human thymus cells. *J. Immunol.* **136**, 1752-1758.

Amiot, M., Dastot, H., Degos, L., Dausset, J., Bernard, A. and Boumsell, L. (1988). HLA class I molecules are associated with CD1a heavy chains on normal human thymus cells. *Proc. Natl.*

Acad. Sci. USA **85**, 4451-4454.

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

Asarnow, D.M., Kuziel, W.A., Bonyhadi, M., Tigelaar, R.E., Tucker, P.W. and Allison, J.P. (1988). Limited diversity of $\gamma\delta$ antigen receptor genes of Thy1$^+$ dendritic epidermal cells. *Cell* **55**, 837-847.

Asarnow, D.M., Goodman, T., Lefrancois, L. and Alison, J.P. (1989). Distinct antigen receptor repertoires of two classes of murine epithelium-associated T cells. *Nature* **341**, 60-62.

Ashton-Rickardt, P.G., Bandeira, A., Delaney, J.R., van Kaer, L., Pircher, H-P., Zinkernagel, R.M. and Tonegawa, S. (1994). Evidence for a differential avidity model of T cell selection in the thymus. *Cell* **76**, 651-663.

Attaya, M., Jameson, S., Martinez, C.K., Hermel, E., Aldrich C., Forman, J., Fischer Lindahl, K., Bevan, M.J. and Monaco, J.J. (1992). Ham-2 corrects the class I antigen prcessing defect in RMA-S cells. *Nature* **354**, 647-649.

Azuma, M., Ito, D., Yagita, H., Okumura, K., Phillips, J.H., Lanier, L.L. and Somoza, C. (1993). B70 antigen is a second ligand for CTLA-4 and CD28. *Nature* **366**, 76-79.

Bahram, S., Bresnahan, M., Geraghty, D.E. and Spies, T. (1994). A second lineage of mammalian major histocompatibility complex class I genes. *Proc. Natl. Acad. Sci. USA.* **91**, 6259-6263.

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

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

Balk, S.P., Ebert, E.C., Blumenthal, R.L., McDermott, F.V., Wucherpfennig, K.W., Landau, S.B. and Blumberg, R.S. (1991b). Oligoclonal expansion and CD1 recognition by human intestinal intraepithelial lymphocytes. *Science* **253**, 1412-1415.

Balk, S.P., Burke, S. Polischuk, J.E., Frantz, M.E., Young, L., Porcelli, S., Colgan, S.P. and Blumberg, R.S. (1994). β_2-microglobulin- independent MHC class Ib molecule expressed by human intestinal epithelium. *Science* **265**, 259-262.

Benoist, C. and Mathis, D. (1989). Positive selection of the T cell repertoire: where and when does it occur. *Cell* **58**, 1027-1033.

Bernabeu, C., van de Rijn, M., Lerch, P. G. and Terhorst, C. P. (1984). β_2-microglobulin from serum associates with MHC class I antigens on the surface of cultured cells. *Nature* **308**, 642-645.

Bikoff, E.K., Huang, L-Y., Episkopou, V., van Meerwijk, J., Germain, R.N. and Robertson, E.J. (1993). Defective major histocompatibility complex class II assembley, transport, peptide acquisition and $CD4^{+}$T cell selection in mice lacking invariant chain expression. *J. Exp. Med.* **177**, 1699-1712.

Bilsland, C.A.G. and Milstein, C. The identification of the β_2-m binding antigen encoded by the human CD1D gene. *Eur. J. Immunol.* **21**, 71-78.

Birnboim, H.C. and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucl. Acids Res.* **7** , 1513-1523.

Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L. and Wiley, D.C.

(1987). Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* **329**, 506-512.

Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L. and Wiley, D.C. (1987). The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature* **329**, 512.

Bjorkman, P.J. and Parham, P. (1990). Structure, function and diversity of class I major histocompatibility complex molecules. *Ann. Rev. Biochem.* **59**, 253-288.

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

Blue, M.-L., Daley, J.F., Levine, H., Craig, K.I. and Schlossman, S.F. (1987). Identification and isolation of a T4+T8+ cell with high T3 expression in human thymus: a possible late intermediate in thymocyte differentiation. *J. Immunol.* **139**, 1065-1069.

Blue, M.-L., Levine, H., Daley, J.F., Branton Jr., K.R. and Schlossman, S.F. (1989). Expression of CD1 and class I MHC antigens by human thymocytes. *J. Immunol.* **141**, 2714-2720.

Bluestone, J.A., Cron, R.Q., Cotterman, M., Houlden, B.A. and Matis, L.A. (1988). Structure and specificity of T cell receptor γ/δ on major histocompatibility complex antigen-specific CD3+, CD4-, CD8- T lymphocytes. *J. Exp. Med.* **168**, 1899-1916.

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

Bonneville, M., Ito, K., Krecko, E.G., Itohara, S., Kappes, D., Ishida, I., Kanagawa, O., Janeway

Bradbury, A. R. M. (1987) PhD Thesis.
University of Cambridge.

Jr., C.J., Murphy, D.B. and Tonegawa, S. (1989). Recognition of a self major histocompatability complex TL region product by $\gamma\delta$ T-cell receptors. *Proc. Natl. Acad. Sci. USA* **86**, 5928-5932.

Born, W., Happ, M.P., Dallas, A., Reardon, C., Kubo, R., Shinnick, T., Brennan, P. and O'Brien, R. (1990). Recognition of heat shock proteins and $\gamma\delta$ T cell function. *Imm. Today* **11**, 40-43.

Boussiotis, V.A., Freeman, G.J., Gribben, J.G., Daley, J., Gray, G. and Nadler, L.M. (1993). Activated human B lymphocytes express three CTLA-4 counterreceptors that stimulate T-cell activation. *Proc. Natl. Acad. Sci. USA.* **90**, 11059-11063.

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

Bradbury, A., Calabi, F. and Milstein, C. (1990). Expression of CD1 in the mouse thymus. *Eur. J. Immunol.* **20**, 1831-1836.

Bradbury, A., Milstein, C. and Kozak, C.A. (1991). Chromosomal localisation of CD1d genes in the mouse. *Somat. Cell Mol. Genet.* **17**, 93-96.

Breitbart, R.E., Andreadis, A. and Nadal-Ginard, B. (1987). Alternative splicing: a ubiquitous mechanism for the generation of multiple protein isoforms from single genes. *Ann. Rev. Biochem.* **56**, 467-495.

Brenner, M.B., McLean, J., Dialynas, D., Strominger, J., Smith, J.A., Owen, F.L., Seidman, J., Ip, S., Rosen, F. and Krangel, M. (1986). Identification of a putative second T cell receptor. *Nature* **322**, 145-149.

Bretscher, A.C. and Cohn, M.M. (1970). A theory of self-nonself discrimination. *Science* **169**, 1042-1049.

Brown, M.G., Driscoll, J and Monaco, J.J. (1991). Structural and serological similarity of MHC-linked LMP and proteasome (multicatalytic proteinase) complexes. *Nature* **353**, 355-357.

Brown, J.H., Jardetsky, T.S., Gorga, J.C., Stern, L.J., Urban, R.G., Strominger, J.L. and Wiley, D.C. (1993). Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* **364**, 33-39.

Bujdoso, R., Hopkins, J., Dutia, B.M., Young, P. and McConnell, I. (1989). Characterisation of sheep afferent lymph dendritic cells and their role in antigen carriage. *J. Exp. Med.* **170**, 1285-1302.

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

Calabi, F., Jarvis, J.M., Martin, L. and Milstein, C. (1989a). Two classes of CD1 genes. *Eur. J. Immunol.* **19**, 285-292.

Calabi, F., Tertia Belt, K., Yung Yu, C., Bradbury, A., Mandy, W.J. and Milstein, C. (1989b). The rabbit CD1 and the evolutionary conservation of the CD1 gene family. *Immunogen.* **30**, 370-377.

Calabi, F., Bilsland, C. A. G., Yung Yu, C., Bradbury,A., Tertia Belt, K., Martin, L. H. and Milstein, C. (1989c). Recent progress in the molecular study of CD1. Pg. 254-258 in *"Leukocyte Typing IV"*. Ed. by W. Knapp et al. Oxford University Press.

Calabi, F., Yu, C.Y., Bilsland, C.A.G. and Milstein, C. (1991). CD1: from structure to function. Chpt.8, Pg.215-243. In *"Immunogenetics of the Major Histocompatibility Complex"*, Ed. by Srivastava, R., Ram, B.P. and Tyle, P. VCH Publishers, Cambridge, England.

Calabi, F. and Bradbury, A. (1991). The CD1 system. *Tissue Antigens.* **37**, 1-9.

Campbell, R.D. and Trowsdale, J. (1993). Map of the human MHC. *Imm. Today* **14**, 349-352.

Carey, N. and Dalziel, R.G. (1994). Sequence variation in the gp135 gene of Maedi visna virus strain EV1. *Virus Genes* **8**, 115-123.

Cattoretti, G., Berti, E., Manusco, A., D'Amato, L., Schiro, R., Soligo, D. and Delia, D. (1988). An MHC class I related family of antigens with widespread distribution on resting and activated cells. Pg. 89-91, In *"Leukocyte Typing III"*. Ed. by A. J. McMichael et al. Oxford Univ Press.

Chan, A.C., Irving, B.A. and Weiss, A. (1992). New insights into T-cell antigen receptor structure and signal transduction. *Curr. Op. in Immunol.* **4**, 246-251.

Chan, A.C., Descai, D.M. and Weiss, A. (1994). The role of protein tyrosine kinases and protein tyrosine phosphatases in T cell antigen receptor signal transduction. *Ann. Rev. Immunol.* **12**, 555-592.

Chan, S.H., Cosgrove, D., Waltzinger, C., Benoist, C. and Mathis, D. (1993). Another view of the selective model of thymocyte development. *Cell* **73**, 225-236.

Chien, Y., Iwashima, M., Kaplan, K.B., Elliott, J.F. and Davis, M.M. (1987). A new T cell receptor gene located within the α locus and expressed in T cell differentiation. *Nature* **327**, 677-682.

Chomczynski, P. and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate- phenol/chloroform extraction. *Anal. Biochem.* **162**, 156-159.

Chorney, M.J., Mashimo, H., Chorney, K.A. and Vasavada, H. (1991). TL genes and antigens.Chpt. 7, Pg. 177-214. In *"Immunogenetics of the Major Histocompatibility Complex"*. Ed. by Srivastava, R., Ram, B.P. and Tyle, P. VCH Publishers, Cambridge, England.

Clevers, H., Alarcon, B., Wileman, T. and Terhorst, C. (1988). The T cell receptor/CD3 complex: a dynamic protein ensemble. *Ann. Rev. Immunol.* **6**, 629-662.

Clontech Laboratories Inc. (1991). RT-PCR- Methods and Applications, Book 1.

Cohen, J. (1993). New protein steals the show as "costimulator" of T cells. *Science* **262**, 844-845.

Connolly, D.J., Cotterill, L.A., Hederer, R.A., Thorpe, C.J., Travers, P.J., McVey, J.H., Dyson, P.J. and Robinson, P.J. (1993). A cDNA clone encoding the mouse Qa-1[a] histocompatibility antigen and proposed structure of the putative peptide binding site. *J. Immunol.* **151**, 6089-6098.

Cotner, T., Mashimo. H., Kung, P.C., Goldstein, G. and Strominger, J. (1981). Human T cell surface antigens bearing a structural relationship to HLA antigens. *Proc. Natl. Acad. Sci. USA* **78**, 3858-3862.

Cresswell, P. (1994). Assembly, transport and function of MHC class II molecules. *Ann. Rev. Immunol.* **12**, 645-674.

Cullen, M.K., Lapiere, L.A., Kesari, K.V. and Geliebter J. (1993). Identification of a recombinogenic major histocompatibility complex Q gene with diverse alleles. *J. Exp. Med.* **177**, 1803-1807.

D'Aquila, R. T., Bechtel, L. J., Videler, J. A., Eron, J. J., Gorczyca, P and Kaplan, J. C. (1991). Maximising sensitivity and specificity of PCR by preamplification heating. *Nucl. Acids. Res.* **19**, 3749.

Davis, M.M. and Bjorkman, P.J. (1988). T cell antigen receptor genes and T cell recognition. *Nature* **334**, 395-402.

Davis, C.B., Killen, N., Caey Crooks, M.E., Raulet, D and Littman, D.R. (1993). Evidence for a stochastic mechanism in the differentiation of mature subsets of T lymphocytes. *Cell* **73**, 237-247.

De Libero, G., Casorati, G., Giachino, C., Carbonara, C., Migone, N., Matzinger, P. and Lanzavecchia, A. (1991). Selection by two powerful antigens may account for the presence of the major population of human peripheral $\gamma\delta$ T cells. *J. Exp. Med.* **173**, 1311-1322.

Dellabona, P., Casorati, G., Friedli, B., Angman, L., Sallusto, F., Tunnacliffe, A., Roosneek, E. and Lanzavecchia, A. (1993). *In vivo* persistence of expanded clones specific for bacterial antigens within the human T cell receptor α/β $CD4^-CD8^-$ subset. *J. Exp. Med.* **177**, 1763-1771.

Dembic, Z., Haas, W., Weiss, S., McCubrey, J., Keifer, H., von Boehmer, H. and Steinmetz, M. (1986). Transfer of specificity by α and β T cell receptor genes. *Nature* **320**, 232-238.

Devereux, J., Haeberli, P. and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the Vax. *Nucl. Acids Res.* **12**, 387-395.

Deverson, E.V., Gow, I.R., Coadwell, W.J., Monaco, J.J., Butcher, G.W. and Howard, J.C. (1990). MHC class II region encoding proteins related to the multi drug resistance family of membrane transporters. *Nature* **348**, 738-741.

Del Porto, P., Mami-Chouib, F., Bruneau, J-M., Jitsukawa, S., Dumas, J., Harnois, M. and Hercend, T. (1991). TCT.1, a target molecule for γ/δ T cells, is encoded by an immunoglobulin superfamily gene (blast-1) located in the CD1 region of human chromosome 1. *J. Exp. Med.* **173**, 1339-1344.

Dent, A.L., Matis, L.A., Hooshmand, F., Widacki, S.M., Bluestone, J.A. and Hedrick, S.M. (1990). Self-reactive $\gamma\delta$ T cells are eliminated in the thymus. *Nature* **343**, 714-719.

Driscoll, J., Brown, M.G., Finley, D. and Monaco, J.J. (1993). MHC-linked LMP gene products alter peptidase activities of the proteasome. *Nature* **365**, 262-264.

Dutia, B.M. and Hopkins, J. (1991). Analysis of the CD1 cluster in sheep. *Vet. Immunol. Immunopath.* **27**, 189-194.

Eckert, K.A. and Kunkel, T.A. (1991). DNA polymerase fidelity and the polymerase chain reaction. *PCR Methods and Applications* **1**, 17-24.

Eghtesady, P., Brorson, K.A., Cheroutre, H., Tigelaar, R.E., Hood, L. and Cronenberg, M. (1992). Expression of mouse Tla region class 1 genes in tissues enriched for γδ T cells. *Immunogen.* **36**, 377-388.

Elliott, J.F., Rock, E.P., Patten, P.A., Davis, M.M. and Chien, Y. (1988). The adult T cell receptor δ chain is diverse and distinct from that of foetal thymocytes. *Nature* **331**, 627-631.

Ennis, P.D., Jackson, A.P. and Parham, P. (1988). Molecular cloning of bovine class I MHC cDNA. *J. Immunol.* **141**, 642-651.

Ennis, P.D., Zemmour, J., Salter, R.D. and Parham, P. (1990). Rapid cloning of HLA-A,B cDNA by using the polymerase chain reaction: frequency and nature of errors produced in amplification. *Proc. Natl. Acad. Sci. USA* **87**, 2833-2837.

Enrietto, P.J., Payne, L.N. and Hayman, M.J. (1983). A recovered avian myelocytomatosis virus that induces lymphomas in chickens. Pathogenic properties and their molecular basis. *Cell* **35**, 369-379.

Falk, K., Rotzschke, O., Stevanovic, S., Jung, G. and Rammensee, H-G. (1991). Allele-specific motifs revealed by sequencing of self peptides eluted from MHC molecules. *Nature* **351**, 290-296.

Faure, F., Jitsukawa, S., Miossec, C. and Hercend, T. (1990). CD1c as a target recognition structure for human T lymphocytes: analysis with peripheral blood γ/δ cells. *Eur. J. Immunol.* **20**, 703-706.

Feinberg, A.P. and Vogelstein, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**, 6

Feinberg, A.P. and Vogelstein, B. (1984). Addendum: A technique for radiolabelling restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **137**, 266

Feitell, M., Mescher, M.F., Olson, C.A., Littman, D.R. and Kronenberg, M. (1991). The thymus leukemia antigen binds human and mouse CD8. *J. Exp. Med.* **174**, 1131-1138.

Feng, D.F. and Doolittle, R. F. (1987). Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Molec. Evol.* **35**, 351-360.

Fischer Lindahl, K., Hermel, E., Loveland, B.E. and Wang, C.R. (1991). Maternally transmitted antigen of mice: a model transplantation antigen. *Ann. Rev. Immunol.* **9**, 351-372.

Flajnik, M.F., Canel, C., Kramer, J. and Kasahara, M. (1991). Evolution of the major histocompatibility complex: molecular cloning of major histocompatibility complex class I from the amphibian *Xenopus*. *Proc. Natl. Acad. Sci. USA.* **88**, 537-541.

Flajnik, M.A., Kasahara, M., Shum, B.P., Salter-Kid, L., Taylor, E. and Du Pasquier, L. (1993). A novel type of gene organisation in vertebrates: a large family of non-MHC-linked class I genes is expressed at the RNA level in the amphibian Xenopus. *EMBO J.* **12**, 4385-4396.

Flavell, R.A., Allen, H., Burkly, L.C., Sherman, D.H., Wareck, G.L. and Widera, G. (1986). Molecular biology of the H-2 histocompatibility complex. *Science* **233**, 437-443.

Germain, R.N. (1994). MHC-dependent antigen processing and peptide presentation : providing ligands for T lymphocyte activation. *Cell* **76**, 287-299.

Forquet, F., Danilczyk, U., Lang, Y and Delovitch, T.L. (1993). Interactions between peptides and major histocompatibility complex molecules during antigen processing and presentation. *Chem. Immunol.* **5 7**, 63-87.

Fraser, J.D., Straus, D. and Weiss, A. (1993). Signal transduction events leading to T-cell lymphokine expression. *Imm. Today* **1 4**, 357-362.

Freeman, G.J., Gribben, J.G., Boussiotis, V.A., Ng, J.W., Restivo, V.A., LOmbard, L.A., Gray, G.S. and Nadler, L.M. (1993a). Cloning of B7-2: a CTLA-4 counter-receptor that costimulates human T cell proliferation. *Science* **2 6 2**, 909-911.

Freeman, G.J., Borriello, F., Hodes, R.J., Reiser, H., Hathcock, K.S., Laszlo, G., McKnight, A.J., Kim, J., Du, L., Lombard, D.B., Gray, G.S., Nadler, L.M and Sharpe, A.H. (1993b). Uncovering of functional alternative CTLA-4 counter-receptor in B7-deficient mice. *Science* **2 6 2**, 907-909.

Fremont, D.H., Matsumara, M., Stura, E.A., Peterson, P. and Wilson, I.A. (1992). Crystal structures of two viral peptides in complex with murine MHC class I H-2K^b. *Science* **2 5 7**, 919-927.

Gambon, F., Kreisler, M. and Diaz-Espada, F. (1988). Correlated expression of surface antigens in human thymocytes. Evidence of class I HLA modulation in thymic maturation. *Eur. J. Immunol.* **1 8**, 153-159.

Garrett, T.P.J., Saper, M.A., Bjorkman, P.J., Strominger, J.L. and Wiley, D.C. (1989). Specificity pockets for the side chains of peptide antigens in HLA-Aw68. *Nature* **3 4 2**, 692-696.

Groettrup, M. and von Boehmer, H. (1993). A role for a pre-T cell receptor in T cell development. *Imm. Today* **1 4**, 610-614.

Groettrup, M., Urgewiss, K., Azogui, O., Palacios, R., Owen, M.J., Hayday, A.C. and von Boehmer, H. (1993). A novel disulphide-linked heterodimer on pre-T cells consists of the T cell receptor β chain and a 33kD glycoprotein. *Cell* **7 5**, 283-294.

Grossbellard, M., Oudet, P. and Chambon, P. (1973). Isolation of high molecular weight DNA from mammalian cells. *Eur. J. Immunol.* **3 6**, 32-38.

Grossberger, D., Hein, W.R. and Marcuz, A. (1990). Class I major histocompatibility complex cDNA clones from sheep thymus: alternative splicing could make a long cytoplasmic tail. *Immunogen.* **3 2**, 77-87.

Guillemot, F., Auffray, C., Orr, H.T. and Strominger, J.L. (1988). Chpt.3. Pg. 81-143. MHC antigens and genes. In "*Molecular Imunology*". Ed. by Hames, B.D. and Glover, D.M. IRL Press.

Haas, W., Pereira, P. and Tonegawa, S. (1993). Gamma/delta T cells. *Ann. Rev. Immunol.* **1 1**, 637-685.

Hata, S., Brenner, M.B. and Krangel, M.S. (1987). Identification of putative human T cell receptor δ cDNA clones. *Science* **2 3 8**, 678-682.

Hathcock, K.S., Laszlo, G., Dickler, H.B., Bradshaw, J., Linsley, P. and Hodes, R.J. (1993). Identification of an alternative CTLA-4 ligand costimulatory for T cell activation. *Science* **2 6 2**, 905-907.

Hanahan, D. (1983). Studies on transformation of *Eschericia coli* with plasmids. *J.Mol. Biol.* **1 6 6**, 557-580.

Hanau, D., Schmitt, D.A., Bieber, T., Schmitt, D. and Cazenave, J.P. (1990). Possible mechanism of action of CD1a antigens. *J. Invest. Dermatol.* **9 5**, 503-505.

Herman, A., Kappler, J.W., Marrack, P. and Pullen, A. M. (1991). Superantigens: mechanism of T cell stimulation and role in immune responses. *Ann. Rev. Immunol.* **9**, 745-772.

Harvey, R. J. and Darlinson, M. G. (1991). Random primed cDNA synthesis facilitates the isolation of multiple 5'-cDNA ends by RACE. *Nucl. Acids Res.* **19**, 4002.

Hashimoto, K., Nakanishi, T. and Kurosawa, Y. (1992). Identification of a shark sequence resembling the major histocompatibility complex class I $\alpha3$ domain. *Proc. Natl. Acad. Sci. USA.* **89**, 2209-2212.

Hedrick, S.M., Cohen, D.I., Nielson, E.A. and Davis, M.M. (1984a). Isolation of cDNA clones encoding T cell specific membrane associated proteins. *Nature* **308**, 149-53.

Hedrick, S.M., Nielson, E.A., Kavaler, J., Cohen, D.I. and Davis, M.M. (1984b). Sequence relationships between putative T cell receptor polypeptides and immunoglobulins. *Nature* **308**, 153-158.

Hedrick, S.M. (1992). Dawn of the hunt for nonclassical MHC function. *Cell* **70**, 177-180.

Hein, W.R., Dudler, L., Beya, M-F., Marcuz, A. and Grossberger, D. (1989). T cell receptor gene expression in sheep: differential usage of TCR1 in the periphery and thymus. *Eur. J. Immunol.* **19**, 2297-2301.

Hein, W.R.and Mackay, C.R. (1991). Prominence of $\gamma\delta$ T cells in the ruminent immune system. *Imm. Today* **12**, 30-34.

Hein, W.R. and Dudler, L. (1993). Divergent evolution of T cell repertoires: extensive diversity and developmentally regulated expression of the sheep $\gamma\delta$ T cell receptor. *EMBO J.* **12**, 715-724.

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

Hogquist, K.A., Gavin, M.A. and Bevan, M.J. (1993). Positive selection of $CD8^+$ T cells induced by major histocompatibility complex binding peptides in foetal thymic organ culture. *J. Exp. Med.* **177**, 1469-1473.

Hogquist, K.A., Jameson, S.C., Heath, W.R., Howard, J.L., Bevan, M.J. and Carbone, F.R. (1994). T cell receptor antagonist peptides induce positive selection. *Cell* **76**, 17-27.

Hopkins, J., Dutia, B.M., Bujdoso, R. and McConnell, I. (1989). *In vivo* modulation of CD1 and MHC class II expression by sheep afferent lymph dendritic cells. *J. Exp. Med.* **170**, 1303-1318.

Hopkins, J. and Dutia, B.M. (1991). Workshop studies on the ovine CD1 homologue. *Vet. Immunol. Immunopath.* **27**, 97-99.

Houlden, B.A., Matis, L.A., Cron, R.Q., Widacki, S.M., Brown, G.D., Pampeno, C., Meruelo, D. and Bluestone, J.A. (1989). A TCR $\gamma\delta$ cell recognising a novel TL-encoded gene product. *Cold Spring Harbor Symposia on Quantitative Biology* **54**, 45-55.

Houlihan, J.M., Biro, P.A., Fergar-Payne, A., Simpson, K.L. and Holmes, C.H. (1992). Evidence for the expression of non-HLA-A, -B, and -C class I genes in the human liver. *J. Immunol.* **149**, 668-675.

Howard, C.J., Sopp, P., Parsons, K., Bembridge, G.P. and Hall, G. (1993a). *Vet. Immunol. and Immunopath.* **39**, 69-76.

Howard, C.J., Sopp, P., Bembridge, G., Young, J. and Parsons, K. (1993b). *Vet. Immunol. and Immunopath.* **39**, 77-83.

Howard, J.C. and Seelig, A. (1993). Peptides and the proteasome. *Nature* **365**, 211-212.

Ichimiya, S., Kikuchi, K. and Matsura, A. (1994). Structural analysis of the rat homologue of CD1. *J. Immunol.* **152**, 1112-1123.

Hughes, A.L. and Nei, M. (1993). Evolutionary relationships of the classes of major histocompatibility complex genes. *Immunogen.* **37**, 337-346.

Hughes, A. (1991). Evolutionary origins and diversification of the mammalian CD1 antigen genes. *Mol. Biol. Evolut.* **8**, 185-201.

Hunkapiller, T. and Hood, L. (1990). Diversity of the immunoglobulin gene superfamily. *Adv. Immunol.* **44**,1.

Hunt, R.F., Henderson, R.A., Shabanowitz, J., Sakaguchi, K., Michel, H., Sevilir, N., Cox , A.L., Appella, V. and Engelhard, V.H. (1992a). Characterisation of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science* **255**, 1261-1263.

Hunt, D.F., Michel, H., Dickinson, T.A., Shabanowitz, J., Cox, A.L., Sakaguchi, K., Appella, E., Grey, H.M. and Sette, A. (1992b). Peptides presented to the immune system by the murine class II MHC molecule I-A^d. *Science* **256**, 1817-1820.

Huynh, T.V., Young, R.A. and Davis, R.W. (1985). Construction and screening of cDNA libraries in λgt10 and λgt11. In DNA Cloning Volume I: a practical approach. (D.M. Glover, Ed.). IRL Press, Oxford.

Imani, F. and Soloski, M.J. (1991). Heat shock proteins can regulate expression of the TLa region-encoded class 1b molecule Qa-1. *Proc. Natl. Acad. Sci. USA* **88**, 10475-10479.

Ish-Horowicz, D. and Burke, J. F. (1981). Rapid and effective cosmid cloning. *Nucl. Acids Res.* **9**, 1989.

Ishii, T., Walsh, L.J., Seymour, G.J. and Powell, R.N. (1990). Modulation of Langerhans cell surface antigen expression by recombinant cytokines. *J. Oral Pathol. Med.* **19**, 355-359.

Ito, K., van Kaer, L., Bonneville, M., Hsu, S., Murphy, D.B. and Tonegawa, S. (1990). Recognition of the product of a novel MHC TL region gene (27^b) by a mouse γδ T cell receptor. *Cell* **62**, 549-561.

Izquierdo, M. and Cantrell, D.A. (1992). T cell activation. *Trends In Cell Biol.* **2**, 268-271.

Janeway, C.A., Jones, B. and Hayday, A. (1988). Specificity and function of T cells bearing γδ receptors. *Imm. Today* **9**, 73-76.

Janeway, C.A. (1992). The T cell receptor as a multicomponent signalling machine: CD4/CD8 corecptors and CD45 in T cell activation. *Ann. Rev. Immunol.* **10**,645-674.

Jenkins, M.K. (1992). The role of cell division in the induction of clonal anergy. *Imm. Today* **13**, 69-73.

Jenkinson, E.J. and Owen, J.J. (1990). T cell differentiation in thymus organ cultures. *Semin. Immunol.* **2**, 51-58.

Jin, Y.J., Clayton, L.K., Howard, F.D., Koyasu, S., Sieh, M., Steinbrich, R., Tarr, G and Reingerz, E.L. (1990). Molecular cloning of the CD3 η subunit identifies a a CD3 η-related product in thymus derived cells. *Proc. Natl. Acad. Sci. USA.* **87**, 3319-3323.

Joyce, S., Tabaczewski, P., Angeletti, R.H., Nathenson, S.G. and Stroynowski, I. (1994). A nonpolymorphic major histocompatibility complex class Ib molecule binds a large array of diverse self-peptides. *J. Exp. Med.* **179**, 579-588.

June, C.H., Bluestone, J.A., Nadler, L.M. and Thompson, C.B. (1994). The B7 and CD28 receptor families. *Imm. Today* **15**, 321-331.

Kahn-Perles, B., Wietzerbin, J., Caillol, D.H. and Lemonnier F. (1985). Delineation of three

Kisielow, P., Bluthmann, H., Staerz, U. D., Steinmetz, M. and von Boehmer, H. (1988). Tolerance in T cell receptor transgenic mice involves deletion of nonmature $CD4^{+}8^{+}$ thymocytes. *Nature* **333**, 742-746.

subsets of human T antigens (HTA) on Molt-4 cells: serological and regulatory relationship to HLA class I antigens. *J. Immunol.* **134**, 1759-1765.

Kasinrerk, W., Baumruker, T., Majdic, O., Knapp, W. and Stockinger, H. (1993). CD1 molecule expression on human monocytes induced by granulocyte-macrophage colony-stimulating factor. *J. Immunol.* **150**, 579-584.

Kaufman, J., Anderson, R., Avila, D., Engberg, D., Lambris, J., Salomonsen, J., Welinder, K. and Skjodt, K. (1992). Different features of the MHC class I heterodimer have evolved at different rates. *J. Immunol.* **148**, 1532-1546.

Kaufmann, S.H.E. (1990). Heatshock proteins and the immune response. *Imm. Today* **11**, 129-136.

Kefford, R.F., Calabi, F., Fearnley, I.M., Burrone, O.R. and Milstein, C. (1984). Serum β_2-microglobulin binds to a T cell differentiation antigen and increases its expression. *Nature* **308**, 641-642.

Kirisits, M.J., Sawai, H., Kunz, H.J. and Gill, T.J.,III. (1994). Multiple TL-like loci in the grc-G/C region of the rat. *Immunogen.* **39**, 301-315.

Kirszenbaum, M., Moreau, P., Glickman, E., Dausset, J. and Carosella, E. (1994). An alternatively spliced form of HLA-G mRNA in human trophoblasts and evidence for the presence of HLA-G transcripts in adult lymphocytes. *Proc. Natl. Acad. Sci. USA.* **91**, 4209-4213.

Kleijmeer, M.J., Kelly, A., Geuze, H.J., Slot, J.W., Townsend, A. and Trowsdale, J. (1992). Location of MHC encoded transporters in the endoplasmic reticulum and cis Golgi. *Nature* **357**, 342-344.

Klein, J., Zhu, Z., Gutknecht, J., Figueroa, F. and Kasahara, M. (1991). MHC: lessons in evolution. Chpt. 2, Pg. 18-38. In *"Immunogenetics of the Major Histocompatibility Complex"*. Ed. by Srivastava, S., Ram, B.P. and Tyle, P. VCH Publishers, Cambridge, England.

Knowles, R.W. and Bodmer, W.F. (1982). A monoclonal antibody recognising a human thymus leukemia antigen associated with β2-micoglobulin. *Eur. J. Immunol.* **12**, 676-681.

Kolenik, S., Ding, T.G. and Longley, J. (1990). Granulocyte-macrophage colony stimulating factor (GM-CSF) decreases CD1a expression by human Langerhans cells and increases proliferation in the mixed epidermal cell lymphocyte reaction (MELR). *J. Invest. Dermatol.* **95**, 359-361.

Kozono, H., White, J., Clements, J. Marrack, P. and Kappler, J. (1994). Production of soluble MHC class II proteins with covalently bound single peptides. *Nature* **369**, 151.

Kronenberg, M., Siu, G., Hood, H.E. and Shastri, N. (1986). The molecular genetics of the T cell antigen receptor and T cell antigen recognition. *Ann. Rev. Immunol.* **4**, 529-591.

Kurlander, R.J., Shawar, S.M., Brown, M.L. and Rich, R.R. (1992). Specialised role for a murine class1-b MHC molecule in prokaryotic host defences. *Science* **257**, 678-679.

Lacasse, J., and Martin, L.H. (1992). Detection of CD1 mRNA in Paneth cells of the mouse intestine by *in situ* hybridisation. *J. Histochem. and Cytochem.* **40**, 1535.

Lamb, C. and Cresswell, P. (1992). Assembley and transport properties of invariant chain trimers and HLA-DR-invariant chain complexes. *J. Immunol.* **148**, 3478-3482.

Lanier, L.L., Allison, J.P. and Phillips J.H. (1986). Correlation of cell surface antigen expression on human thymocytes by multicolour flow cytometric analysis: implications for differentiation. *J. Immunol.* **137**, 2501-2507.

Lanzavecchia, A. (1993). *In vivo* persistence of expanded clones specific for bacterial antigens within the human T cell receptor α/β CD4$^-$CD8$^-$ subset. *J. Exp. Med.* **177**, 1763-1771.

Latron, F., Pazmany, L., Morrison, J., Moots, R., Saper, M.A., McMichael, A. and Strominger, J.L. (1992). A critical role for conserved residues in the cleft of HLA-A2 in presentation of a nonapeptide to T cells. *Science* **257**, 964-967.

Lawlor, D.A., Zemmour, J., Ennis, P.D. and Parham, P. (1990). Evolution of class I MHC like genes and proteins: from natural selection to thymic selection. *Ann. Rev. Immunol.* **8**, 23-63.

Ledbetter, J.A., Tsu, T.T. and Clark, E.A. (1985). Covalent association between human thymus leukemia-like antigens and CD8 (Tp32) molecules. *J. Immunol.* **134**, 4250-4254.

Lenschow, D.J., Su, G. H-T., Zuckerman, L.A., Nabavi, N., Jellis, C.L., Gray, G.S., Miller, J. and Bluestone, J.A. (1993). Expression and functional significance of an additional ligand for CTLA-4. *Proc. Natl. Acad. Sci. USA.* **90**, 11054-11058.

Lerch, P.G., van de Rijn, M., Schrier, P. and Terhorst, C. (1983). Biochemical comparison of the T6 antigen and HLA-A, B antigens. *Human Immunol.* **6**, 13-30.

Lew, A. M., McCluskey, J., Maloy, W. L., Marguiles, D. H. and Colligan, J. E. (1987). Multiple class I molecules generated from single genes by alternative splicing of pre-mRNA's. *Imunol. Res.* **6**, 117-132.

Linsley, P.S. and Ledbetter, J.A. (1993). The role of the CD28 receptor during T cell responses to antigen. *Ann. Rev. Immunol.* **11**, 191-212.

Loh, E. Y., Elliott, J. F., Cwirla, S., Lanier, L. L. and Davis, M. M. (1989). Polymerase chain reaction with single sided specificity: analysis of T cell receptor δ chain. *Science* **243**, 217-

220.

Loss, G. and Sant, A. (1993). Invariant chain retains MHC class II molecules in the endocytic pathway. *J. Immunol.* **150**, 318.

McCluskey, J., Boyd, J.F., Maly, W.L., Coligan, J.E. and Margulies, D.H. (1987). Multiple class I molecules generated from single genes by alternative splicing of pre-mRNA's. *Immunol. Res.* **6**, 117-132.

MacHugh, N.D., Bensiad, A., Davis, W.C., Howard, C.J., Parsons, K.R., Jones, B. and Kaushal, A. (1988). Characterisation of a bovine thymic differentiation antigen analogous to CD1 in the human. *Scand. J. Immunol.* **27**, 541-547.

MacKay, C.R., Maddox, J.F., Gogolin-Ewens, K.J. and Brandon, M.R. (1985). Characterisation of two sheep lymphocyte differentiation antigens, SBU-T1 and SBU-T6. *Immunol.* **55**, 729-737.

Madden, D.R., Gorga, J.C., Strominger, J.L. and Wiley, D.C. (1991). The structure of HLA-B27 reveals nonamer self-peptides bound in an extended conformation. *Nature* **353**, 321-325.

Madden, D.R., Gorga, J.C., Strominger, J.L. and Wiley, D.C. (1992). The three-dimensional structure of HLA-B27 at 2.1 A resolution suggests a general mechanism for tight peptide binding to MHC. *Cell* **70**, 1035-1048.

Malissen, M., Gillet, A., Rocha, B., Trucy, J., Vivier, V., Boyer, C., Kontgen, F., Brun, N., Mazza, G., Spanopoulou, E., Guy-Grand, D. and Malissen, B. (1993). T cell development in mice lacking the CD3-ζ/η gene. *EMBO J.* **12**, 4347-4355.

Mallick, C.A., Dudley, E.C., Viney, J.L., Owen, M.J. and Hayday, A.C. (1993). Rearrangement and diversity of T cell receptor β chain genes in thymocytes: a critical role for the β chain in

Mami-Chouaib F., del Porto, P., Delorme, D. and Hercend, T. (1991). Further evidence for a $\gamma\delta$ T cell receptor-mediated TCT.1/CD48 recognition. *J. Immunol.* **147**, 2864-2867.

development. *Cell* **73**, 513-519.

Mami-Chouib, F., Miossec, C., Del Porto, P., Flament, C., Friebel, F. and Hercend, T. (1990). T cell target 1 (TCT.1): a novel target molecule for human non-major histocompatibility complex restricted T lymphocytes. *J. Exp. Med.* **172**, 1071-1082.

Mann, D. and Forman, J. (1988). Recognition by cytotoxic T lymphocytes of Qa-2 antigens. *J. Immunol.* **141**, 1813-1818.

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

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

Matis, L.M. and Bluestone, J.A. (1991). Specificity of $\gamma\delta$ receptor bearing T cells. *Semin. Immunol.* **3**, 75-80.

Matsumara, M., Fremont, D.H., Peterson, P. and Wilson, I.A. (1992). Emerging principles for the recognition of peptide antigens by MHC class I molecules. *Science* **257**, 927-934.

Meuer, S., Fitzgerald, K., Hossey, H., Hodgdon, J., Schlossman, S. and Reinherz, E. (1983). Clonotypic structures involved in antigen specific human T cell function. *J. Exp. Med.* **157**, 705-719.

Milligan, G.N., Flaherty, L., Braciele, V.L. and Braciele, T.J. (1991). Nonconventional (TL-encoded) major histocompatibility complex molecules present processed viral antigen to cytotoxic T lymphocytes. *J. Exp. Med.* **174**, 133-138.

Momburg, F., Neefjes, J. and Hammerling, G.J. (1994). Peptide selection by MHC encoded TAP transporters. *Curr. Op. Immunol.* **6,** 32-37.

Morita, T., Takahashi, T., Stockert, E., Nakayama, E., Tsuji, T., Matsudaira, Y., Old, L. J. and Obata, Y. (1994). TL antigen as a transplantation antigen recognised by TL-restricted cytotoxic T cells. *J. Exp. Med.* **179**, 777-784.

Milstein, C., Calabi, F., Jarvis, J.M., Kefford, R., Martin, L. and Migone, N. (1987). CD1: a family of MHC-related genes which do not map in chromosome 6. Pg. 882-889. In *"Leukocyte Typing III"*. Ed. by A.J. McMichael et al. Oxford Univ. Press.

Monaco, J.J. (1992). A molecular model of MHC class I-restricted antigen processing. *Imm. Today* **1 3**, 173-184.

Moodie, S.A. and Wolfman, A. (1994). The 3Rs of life: Ras, Raf and growth regulation. *TIGS* **1 0**, 44-48.

Moretta, L., Ciccone, E., Ferrine, S., Pelicci, P.G., Mingari, M.C., Zenomski, J., Bottino, C., Grossi, C. and Moretta, A. (1991). Molecular and cellular analysis of human T lymphocytes expressing the $\gamma\delta$ T cell receptor. *Imm. Rev.* **1 2 0**, 117-135.

Moriwaki, S., Korn, B.S., Ichokawa, Y., van Kaer, L. and Tonegawa, S. (1993). Amino acid substitutions in the floor of the putative antigen binding site of H-2T22 affect recognition by a $\gamma\delta$ T-cell receptor. *Proc. Natl. Acad. Sci. USA.* **9 0**, 11396-11400.

Morse, R.Y., Watta, S., Gelber, C. and Goodenow, R.S. (1991). The Qa region genes and antigens of the murine major histocompatibility complex. Chpt.6, Pg. 155-176. In *"Immunogenetics of the Major Histocompatibility Complex"*. Ed. by Srivastava, S., Ram, B.P. and Tyle, P. VCH Publishers, Cambridge, England.

Mosser, D.D., Duchaine, J. and Martin, L.H. (1991). Biochemical and developmental characterisation of the murine cluster of differentiation 1 antigen. *Immunol.* **7 3**, 298-303.

Mossman, T.R. and Coffman, R.L. (1989). TH1 and TH2 cells: different patterns of lymphokine secretion lead to differential properties. *Ann. Rev. Immunol.* **7** , 145.

Mullis, K. B. and Faloona, F. A. (1987). Specific synthesis of DNA *in vitro* via a polymerase

catalysed chain reaction. *Methods Enzymol.* **155**, 335-350.

Needleman, S. B. and Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**, 443-453.

Neefjes, J.J., Momburg, F. and Hammerling, G.J. (1993). Selective and ATP-dependent translocation of peptides by the MHC-encoded transporter. Science 261, 769-771.

Nikolic-Zugic, J. and Bevan, M.J. (1990). Role of self-peptides in positively selecting the T cell repertoire. *Nature* **344**, 65-67.

Nossal G.J.V. (1994). Negative selection of lymphocytes. *Cell* **76**, 229-239.

Ohara, O., Dorit, R. L. and Gilbert, W. (1989). One-sided polymerase chain reaction: the amplification of cDNA. *Proc. Natl. Acad. Sci. USA.* **86**, 5673-5677.

Ohno, H., Aoe, T., Taki, S., Kitamura, D., Ishida, Y., Rajewsky, K. and Saito, T. (1993). *EMBO J.* **12**, 4357-4366.

Padovan, E., Casorati, G., Dellabona, P., Meyer, S., Brockhaus, M. and Lanzavecchia, A. (1993). Frequent expression of two TCR α chains on human $\alpha\beta$ lymphocytes creates dual receptor cells. *Science* **262**, 422-424.

Pamer, E.G., Wang, C.R., Flaherty, L., Fischer-Lindahl, K. and Bevan, M. (1992). H-2M3 presents a *Listeria monocytogenes* peptide to cytotoxic T lymphocytes. *Cell* **70**, 215-223.

Panja, A., Blumberg, R.S., Balk, S.P. and Mayer, L. (1993). CD1d is involved in T cell-intestinal epithelial cell interactions. *J. Exp. Med.* **178**, 1115-1119.

Parham, P. (1992). Deconstructing the MHC. *Nature* **360**, 300-301.

Pescovitz, M.D., Hsu,.S-M., Katz, S.I., Lunney, J.K., Shimada.S., Sachs,.D,H. (1990). Characterisation of a porcine CD1-specific mAb that distinguishes CD4/CD8 double-positive thymic from peripheral T lymphocytes. *Tissue Antigens* **35**, 151-156.

Ploegh, H. and Benaroch, P. (1993). MHC class II dimer of dimers. *Nature* **364**, 16-17.

Pomp, D. and Medrano, J. F. (1991). Organic solvents as facilitators of polymerase chain reactions. *Biotechniques* **10**, 58-59.

Porcelli, S., Brenner, M.B., Greenstein, J.L., Balk, S.P., Terhorst, C. and Bleicher, P.A. (1989). Recognition of cluster of differentiation 1 antigens by human $CD4^{-}8^{-}$ cytolytic T lymphocytes. *Nature* **341**, 447-450.

Porcelli, S., Brenner, M. and Band, H. (1991). Biology of the human $\gamma\delta$ T cell receptor. *Immunol. Rev.* **120**, 137-183.

Porcelli, S., Morita, C.T. and Brenner, M.B. (1992). CD1b restricts the response of human $CD4^{-}8^{-}$ T lymphocytes to a microbial antigen. *Nature* **360**, 593-597.

Potter, T.A., Rajan, T.V., Dick, R.F. II and Bluestone, J.A. (1989). Substitution at position 227 of H-2 class I molecules abrogates recognition by CD8-dependent, but not CD8-independent, cytotoxic T lymphocytes. *Nature* **337**, 73-75.

Powis, S.J., Deverson, E.V., Coadwell, W.J., Ciruela, A., Huskisson, N.S., Smith, H., Butcher, G.W. and Howard, J.C. (1992a). Effect of polymorphism of an MHC-linked transporter on the peptides assembled in a class I molecule. *Nature* **357**, 211-215.

Powis, S.H., Mockridge, I., Kelly, A., Kerr, L-A., Glynne, R., Gileadi, U., Beck, S. and Trowsdale, J. (1992b). Polymorphism in a second ABC transporter gene located within the class II region of the human major histocompatibility complex. *Proc. Natl. Acad. Sci. USA.* **89**,

Rammensee, H. G. , Falk, K and Rotzschke, O. (1993a). Peptides naturally associated with MHC class I molecules. *Ann. Rev. Immunol.* **11**, 213-244.

1463-1467.

Rammensee, H-G., Rotzsche, O. and Falk, K. (1993b). MHC class I-restricted antigen processing- lessons from natural ligands. *Chem. Immunol.* **57**, 113-133.

Raulet, D.H., Garman, R.D., Saito, H. and Tonegawa, S. (1985). Developmental regulation of T cell receptor gene expression. *Nature* **314**, 103-107.

Raulet, D.H. (1989a). The structure, function, and molecular genetics of the γδ T cell receptor. *Ann. Rev. Immunol.* **7**, 175-207.

Raulet, D.H. (1989b). Antigens for γδ T cells. *Nature* **339**, 342-343.

Razi-Wolf, Z., Galvin, F., Gray, G. and Reiser, H. (1993). Evidence for an additional ligand, distinct from B7, for the CTLA-4 receptor. *Proc. Natl. Acad. Sci. USA.* **90**, 11182-11186.

Reinherz, E.L., Kung, P.C., Goldstein, G., Levey, R. and Schlossman. (1980). Discrete stages of human intrathymic differentiation: analysis of normal thymocytes and leukemic lymphoblasts of T cell lineage. *Proc. Natl. Acad. Sci. USA* **77**, 1588-1592.

Riberdy, J.M., Newcomb, J.R., Surman, M.J., Barbosa, J.A. and Cresswell, P. (1992). HLA-DR molecules from an antigen processing cell line are associated with invariant chain peptides. *Nature* **360**, 474-477.

Rippman, F., Taylor, W.R., Rothbard, J.B. and Green, N.M. (1991). A hypothetical model for the peptide binding domain of hsp70 based on the peptide binding domain of HLA. *EMBO J.* **10**, 1053-1059.

Robey, E.A., Fowlkes, B., Gordon, J.W., Kioussis, D., von Boehmer, H., Ramsdell, F. and Axel, R. (1991). Thymic selection in CD8 transgenic mice supports an instructive model for

committment to a CD4 or CD8 lineage. *Cell* **64**, 99-107.

Robey, E. and Fowlkes, B.J. (1994). Selective events in T cell development. *Ann. Rev. Immunol.* **12**, 675-705.

Roche, P. and Cresswell, P. (1990). Invariant chain association with HLA-DR molecules inhibits immunogenic peptide binding. *Nature* **345**, 615-618.

Rock, E.P., Sibbald, P.R., Davis, M.M. and Chien, Y-H. (1994). CDR3 length in antigen specific immune receptors. *J. Exp. Med.* **179**, 985-989.

Rogers, J. J., Savarski, D. F., Shacter, E., Maloy, W. L., Lillehoj, E. P. and Coligan, J. E. (1986). Three distinct H-2K^s molecules differing at the carboxyl terminus are expressed on a tumor from SJL/J mice. *J. Immunol.* **137**, 3006-3011.

Rolfs, A., Schuller, I., Finckh, U. and Weber-Rolfs, I. (1992). *"PCR: clinical diagnostics and research"*. Pub. by Springer-Verlag.

Rothenberg, E.V. (1992). The development of functionally responsive T cells. *Adv. Immunol.* **51**, 85-214.

Rothermel, E., Heine, L., Wurst, W. and Gunther, E. (1993). Characterisation of a class Ib gene of the rat major histocompatibility complex. *Immunogen.* **38**, 82-91.

Rotzschke, O., Falk, K., Stevanovic, S., Grahovac, B., Soloski, M.J., Jung, G. and Rammensee, H-G. (1993). Qa-2 molecules are peptide receptors of higher stringency than ordinary class I molecules. *Nature* **361**, 642-644.

Rudensky, A. and Janeay, C.A. (1993). Studies on naturally processed peptides associated with MHC class II molecules. *Chem. Immunol.* **57**, 134-151.

Saito, H., Kranz, D.M., Takagaki, Y., Hayday, H.C., Eisen, H.N. and Tonegawa, S. (1984). Complete primary structure of a heterodimeric T cell receptor deduced from cDNA sequences. *Nature* **309**, 757-762.

Saito, T., Weiss, A., Miller, J., Norcross, M.A. and Germain, R.N. (1987). Specific antigen activation of transfected human cells expressing murine Ti alpha and beta human T3 receptor complexes. *Nature* **325**, 125-130.

Salami, M. A., Simpson-Morgan, M. W. and Morris, B. (1985). Haemopoiesis and the development of immune reactivity in the sheep foetus. Pg. 19-45 in *"Immunology Of The Sheep"*. Ed. by B. Morris and M. Miyasaka.

Salamone, M. del C. and Fainboim L. (1992). Intracellular expression of CD1 molecules on PHA-activated normal T lymphocytes. *Immunol. Letters* 33, 61-66.

Salter, R., Norment, A., Chen, B., Clayberger, C., Krensky, A., Littman, D. and Parham, P. (1989). Polymorphism in the α3 domain of HLA-A molecules affects binding to CD8. *Nature* **338**, 345-347.

Salter, R.D., Benjamin, R.J., Wesley, P.K., Buxton, S.E., Garrett, T.P.J., Clayberger, C., Krensky, A.M., Norment, A.M., Littman, D.R. and Parham, P. (1990). A binding site for the T cell co-receptor CD8 on the α3 domain of HLA-A2. *Nature* **345**, 41-46.

Sambrook, J., Fritsch, E. F. and Maniatis T. (1989). Molecular Cloning- a Laboratory Manual. Second Edition. Cold Spring Harbour Laboratory Press.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.

Saper, M.A., Bjorkman, P.J. and Wiley, D.C. (1991). Refined structure of the human

Sciammas, R., Johnson, R. M., Sperling, A. I., Brady, W., Linsley, P.S., Spear, P.G., Fitch, F.W. and Bluestone, J. A. (1994). Unique antigen recognitionby a herpes virus-specific TCR-$\gamma\delta$ cell. *J. Immunol.* **152**, 5392-5397.

histocompatibility antigen HLA-A2 at 2.6 A resolution. *J.Mol.Biol.* **219**, 277-319.

Schild, H., Mavaddat, N., Litzenberger, C., Ehrich, E.H., Davis, M.M., Bluestone, J.A., Matis, L. Draper, R.K. and Chien, Y.-H. (1994). The nature of major histocompatibility complex recognition by $\gamma\delta$ T cells. *Cell* **76**, 29-37.

Schmid, S.L. and Jackson, M.R. (1994). Making class II presentable. *Nature* **369**, 103-104.

Scholl, P.R and Geha, R.S. (1994). MHC class II signalling in B cell activation. *Imm. Today* **15**, 418-422.

Shawar, S.M., Cook, R.D., Rodgers, J.R. and Rich, R.R. (1990). Specialised functions of MHC Class I molecules. I. An N-formyl peptide receptor is required for construction of the class I antigen Mta. *J. Exp. Med.* **171**, 897-912.

Shawar, S.M., Vyas, J.M., Rodgers, J.R., Cook, R.G. and Rich, R.R. (1991). Specialised functions of major histocompatibility complex class I molecules: II. Hmt binds N-formylated peptides of mitochondrial and prokaryotic origin. *J. Exp. Med.* **174**, 941-944.

Shawar, S.M., Vyas, J.M., Rodgers, J.R. and Rich, R.R. (1994). Antigen presentation by major histocompatibility complex class Ib molecules. *Ann. Rev. Immunol.* **12**, 839-880.

Shull, G.E., Schwartz, A. and Lingel, J.B. (1985). Amino acid sequence of the catalytic subunit of (Na^{+}+ K^{+}) ATPase deduced from a complementary DNA. *Nature* **316**, 691-695.

Shum, B.P., Avila, D., Du Pasquier, L., Kasahara, M and Flajnik, M.F. (1993). Isolation of a classical MHC class I cDNA from an Amphibian. *J. Immunol.* **151**, 5376-5386.

Simpson, E. (1992). Positive and negative selection of the T cell repertoire: role of MHC and other ligands. *Intern. Rev. Immunol.* **8**, 269-277.

Small, T.N., Knowles, R.W., Keever, C., Kernen, N.A., Collins, N., O'Reilly, R.J., Dupont, B. and Flomenberg, N. (1987). M241 (CD1) expression on B lymphocytes. *J. Immunol.* **138**, 2864-2868.

Snodgrass, H.R., Dembic, Z., Steinmetz, M. and von Boehmer, H. (1985). Expression of T cell antigen receptor genes during foetal development in the thymus. *Nature* **315**, 232-233.

Snow, P.M., van de Rijn, M. andTerhorst, C. (1985). Association between the human thymic differentiation antigens T6 and T8. *Immunol.* **15**, 529-532.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503.

Spies, T., Bresnahan, M., Bahram, S., Arnold, D., Blanck, G., Mellins, E., Pious, D. and Demars, R. (1990). A gene in the major histocompatibility class II region controlling the class I antigen presentation pathway. *Nature* **348**, 744-747.

Spies, T. and Demars, R. (1991). Restored expression of MHC class I molecules by gene transfer of a putative peptide transporter. *Nature* **351**, 323-324.

Sprent, J. and Webb, S.R. (1987). Function and specificity of T cell subsets in the mouse. *Adv. Immunol.* **41**, 39-133.

Srivastava, R. and Lambert, M.E. (1991). Molecular organisation and expression of class I HLA gene family. Chpt. 5, Pg. 100-154. In "*Immunogenetics of the Major Histocompatibility Complex*". Ed. by Srivastava, S., Ram. B.P. and Tyle, P. VCH Publishers, Cambridge, England.

Strominger, J. L. (1989). The $\gamma\delta$ T cell receptor and class 1b MHC-related proteins: enigmatic molecules of immune recognition. *Cell* **57**, 895-898.

Stroynowski, I. (1990). Molecules related to class I major histocompatibility antigens. *Ann. Rev. Immunol.* **8**, 501-530.

Stroynowski, I. and Fischer Lindahl, K. (1994). Antigen presentation by nonclassical class I molecules. *Curr. Op. Immunol.* **6**, 38-44.

Tabaczewski, P., Shirwar, H., Lewis, K. and Stroynowski, I. (1994). Alternative splicing of class Ib major histocompatibility complex transcripts *in vivo* lead to the expression of soluble Qa-2 molecules in murine blood. *Proc. Natl. Acad. Sci. USA.* **91**, 1883-1887.

Terhorst, C., Agthoven, A.van, LeClair, K., Snow, P., Reinherz, E. and Schlossman, S. (1980). Biochemical studies of the human thymocyte cell surface antigens T6, T9 and T10. *Cell* **23**, 771-780.

Terhorst, C. and Regueiro, J.R. (1993). T cell activation. Chpt. 24, Pg. 447-466. In "*Clinical Aspects of Immunology*", vol.1. Ed. by Lachmann, P.J., Peters, K., Rosen, F.S. and Walport, M.J. 5th edition. Blackwell Sci. Publications.

Townsend, A., Ohlen, C., Bastin, J., Ljunggren, H-G., Foster, L. and Karre, K. (1989). Association of class I major histocompatibility heavy and light chains induced by viral peptides. *Nature* **340**, 443-448.

Theodorou, I.D., Boumsell, L. Calvo, C,-F., Govy, H., Beral, H.M. and Debre, P. (1990). CD1 stimulation of human T cell lines induces a rapid increase in the intracellular free Ca^{2+} concentration and the production of IL-2. *J. Immunol.* **144**, 2518-2523.

Tindall, K. R. and Kunkel, T. A. (1988). Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Biochem.* **27**, 6008-6013.

Toyonaga, B. and Mak, T.W. (1987). T cell antigen receptor in normal and malignant cells.

Ann. Rev. Immunol. **5**, 585-620.

Tranecker, A., Oliveri, F., Allen, N. and Karajalainen, K. (1986). Normal T cell development is possible without functional γ chain genes. *EMBO J.* **5**, 1589-1593.

Trowsdale J., Hanson, I., Mockridge, I., Beck, S., Townsend, A. and Kelly, A. (1990). Sequences encoded in the MHC class II region of the MHC related to the ABC superfamily of transporters. *Nature* **348**, 741-744.

Tulp, A., Verwoerd, D., Dobberstein, B., Ploegh, H.L., Pieters, J. (1994). Isolation and characterisation of the intracellular MHC class II compartment. *Nature* **369**, 120-126.

Tysoe-Calnon, V.A., Grundy, J.E. and Perkins, S.J. (1991). Molecular comparisons of the β_2-microglobulin binding site in class I major histocompatibility complex α chains and proteins of related sequence. *Biochem J.* **277**, 359-369.

Ulbrecht, M., Rehberger, B., Strobel, I., Messer, G., Kind, P., Degitz, K., Bieber, T. and Weiss, E.H. (1994). HLA-G: expression in human keratinocytes *in vitro* and in human skin *in vivo*. *Eur. J. Immunol.* **24**, 176-180.

van Agthoven, A. and Terhorst, C. (1982). Further biochemical characterisation of the human thymocyte differentiation antigen T6. *J. Immunol.* **128**, 426-432.

van Bleek, G.M. and Nathenson, S.G. (1990). Isolation of an endogenously processed immunodominant viral peptide from the class I H-2K^b molecule. *Nature* **348**, 213-216.

van Bleek, G.M and Nathenson, S.G. (1993). Peptide presentation by major histocompatibility class I molecules. *Chem. Immunol.* **57**, 1-17.

Vanbuskirk, A., Cromp, B.L., Margoliash, E. and Pierce, S.K. (1989). A peptide binding

protein having a role in antigen presentation is a member of the HSP70 heat shock family. *J. Exp. Med.* **170**, 1799-1809.

van de Rijn, M., Lerch, P.G., Knowles, R.W. and Terhorst, C. (1983). The thymic differentiation markers T6 and M241 are two unusual MHC class I antigens. *J. Immunol.* **131**, 851-855.

Viville,S., Neefjes, J., Lotteau, V., Dierich, A., Lemeur, M., Ploegh, H., Benoist, C. and Matis, D. (1993). Mice lacking the MHC class II associated invariant chain. *Cell* **72**, 635-648.

von Boehmer, H. (1986). The selection of the $\alpha\beta$ heterodimeric T cell receptor for antigen. *Imm. Today* **7**, 333-336.

von Boehmer, H. and Kisielow, P. (1993). Lymphocyte committment: instruction versus selection. *Cell* **73**, 207-208.

von Boehmer, H. (1994). Positive selection of lymphocytes. *Cell* **76**, 219-228.

Vidovic, D., Roglic, M., McKune, K., Guerder, S., MacKay, C. and Dembic, Z. (1989). Qa-1 restricted recognition of foreign antigen by a $\gamma\delta$ T-cell hybridoma. *Nature* **340**, 646-650.

Vyas, J.M., Shawar, S.M., Rodgers, J.R., Cook, R.G. and Rich, R.R. (1992). Biochemical specificity of H-2M3^{a}: stereospecificity and space-filling requirements at position 1 maintain N-formyl peptide binding. *J. Immunol.* **149**, 3605-3611.

Vyas, J.M., Rich, R.R., Howell, D.D., Shawar, S.M. and Rodgers, J.R. (1994). Availability of endogenous peptides limits expression of an M3a-Ld major histocompatibility complex class I chimera. *J. Exp. Med.* **179**, 155-165.

Yanagi, Y., Yoshikai, Y., Legget, K.,Clark, S.P., Alexander, J. and Mak, T.W. (1984). A human T cell specific cDNA clone encodes a protein having extensive homolgy to immunoglobulin

chains. *Nature* **308**, 145-149.

Yannisch-Perron, C., Viera, J. and Messing, J. (1985). Improved M13 'phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103.

Yewdell, J.W. and Bennink, J.R. (1992). Cell biology of antigen processing and presentation to major histocompatibility complex class I molecule restricted T lymphocytes. *Adv. Immunol.* **52**, 1-123.

Wahle, E. and Keller, W. (1992). The biochemistry of 3' end cleavage and polyadenylation of messenger RNA precursors. *Ann. Rev. Biochem.* **61**, 419-440.

Wang, C.-R., Loveland, B.E. and Fischer Lindahl, K. (1991). H-2M3 encodes the MHC class I molecule presenting the maternally transmitted antigen of the mouse. *Cell* **66**, 335-345.

Wang, C.-R., and Fischer Lindahl, K. (1993a). HMT, encoded by H-2M3, is a neoclassical major histocompatibility class I antigen. *Proc. Natl. Acad. Sci. USA* **90**, 2784-2788.

Wang, Q., Geliebster, J., Tonkonogy, S. and Flaherty, L. (1993b). Expression of the Q2 gene of the MHC in thymus and intestinal epithelial cells. *Immunogen.* **38**, 370-372.

Wang, C-R. and Fischer Lindahl, K. (1993c). Organisation and structure of the H2-M4-M8 class I genes in the mouse major histocompatibility complex. *Immunogen.* **38**, 258-271.

Weiss, A. (1993). T cell antigen receptor signal transduction: a tale of tails and cytoplasmic protein-tyrosine kinases. *Cell* **73**, 209-212.

West, M.A., Lucocq, J.M. and Watts, C. (1994). Antigen processing and class II peptide-loading compartments in human B-lymphoblastoid cells. *Nature* **369**, 147-150.

Wickens, M. (1990). How the messenger got its tail: addition of poly (A) in the nucleus. *TIBS* **15**, 277-281.

Woolfson, A. and Milstein, C. (1994). Alternative splicing generates secretory isoforms of human CD1. *Proc. Natl. Acad. Sci. USA.* **91**, 6683-6687.

Wu, M., van Kaer, L., Itohara, S. and Tonegawa, S. (1991). Highly restricted expression of the thymus leukemia antigens on intestinal epithelium cells. *J. Exp. Med.* **174**, 213-218.

Young, R.A. and Elliott, T.J. (1989). Stress proteins, infection and immune surveillance. *Cell* **59**, 5-8.

Yung Yu, C. and Milstein,C. (1989). A physical map linking the five CD1 human thymocyte differentiation antigens. *EMBO J.* **8**, 3727-3732.

Ziegler, A. and Milstein, C. (1979). A small polypeptide different from β_2-microglobulin associated with a human cell surface antigen. *Nature* **279**, 243-244.

Zinkernagel, R.M. and Doherty, P.C. (1974). Restriction of *in vitro* mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semi-allogeneic system. *Nature* **248**, 701-702.

Zhang, W., Young, A.C.M., Imarai, M., Nathenson, S.G. and Sacchettini, J.C. (1992). Crystal structure of the major histocompatibility complex class I H-2K^b molecule containing a single viral peptide: implications for peptide binding and T cell recognition. *Proc. Natl. Acad. Sci. USA.* **89**, 8403-8407.

Appendix 1

General Buffers

PBS (Phosphate-buffered saline)

137mM NaCl

27mM KCl

8mM Na_2HPO_4

15mM KH_2PO_4

Adjust to pH7.2

10xTE buffer (Tris-EDTA)

100mM Tris.Cl, pH8.0

10mM EDTA, pH8.0

10xTAE buffer (Tris-Acetate-EDTA)

400mM Tris.acetate

10mM EDTA, pH8.0

10xTBE buffer (Tris-borate-EDTA)

900mM Tris.borate

20mM EDTA, pH8.0

Adjust to pH8.3

This buffer is equivalent to 10xTBE-M buffer.

10xTBE-A buffer

13.4mM Tris base

4.5mM boric acid

10mM EDTA, pH8.0

STE (sodium chloride-tris-EDTA)

0.1M NaCl

10mM Tris.Cl pH8.0

1mM EDTA pH8.0

20xSSC buffer (Sodium chloride-sodium citrate)

3M NaCl

300mM Tris.sodium citrate

Adjust to pH7.0

Solutions Used For RNA/DNA Preparations

DEPC-Treated SDW

Sterile deionised distilled water is treated overnight with 0.02% diethylpyrocarbonate (DEPC) and then autoclaved at 15lbs/in^2 form 15 minutes to inactivate the DEPC.

GTC (Guanidinium isothiocyanate)

4M guanidinium isothiocyanate

0.5% (w/v) N-lauryl sarcosine

25mM sodium citrate pH7.0

0.7% (v/v)2-mercaptoethanol

10xMOPS (morpholino-propanesulfonic acid

0.2M 3-(N-morpholino)-propanesulfonic acid

0.05M sodium acetate

0.01M EDTA

41.8g MOPS is added to 800ml DEPC-treated SDW and the pH adjusted to 7.0 with NaOH or acetic acid. Add 16.6ml of 3M DEPC-treated sodium acetate and 20ml 0.5M DEPC-treated EDTA, pH8.0. Make the volume up to 1 litre with DEPC-treated SDW and filter sterilise.

RNA loading buffer

500µl formamide

100µl 10xMOPS

150µl formaldehyde

2µl ethidium bromide@10mg/ml

use at 3-5x sample volume.

RNA dye mix

30% (w/v) ficoll

0.05% (w/v) bromophenol blue

5xMOPS

100μg/ml ethidium bromide

Phenol/Chloroform

mix redistilled phenol and chloroform in a 1:1 ratio.

Add hydroxyquinolone to 0.05% (w/v).

Equilibrate the mixture to pH7.4 with 1xTE or with DEPC-treated SDW for RNA work.

DNA sample buffer

40% (w/v) sucrose

0.25% (w/v) bromophenol blue

0.25% (w/v) xylene cyanol FF

TFB solution (Hanahans buffer)

10mM K-MES pH6.2 (potassium salt of 2N-morpholinoethane sulfonic acid)

100mM RbCl

45mM $MnCl_2.4H_2O$

10mM $CaCl_2.2H_2O$

3mM $HaCoCl_3$ (hexamine cobalt [III]chloride)

Adjust to pH6.5.

Streile filter through 0.2μm filter and store at 4°C.

TfBI

0.035M sodium acetate

0.01M $CaCl_2$

0.1M RbCl

0.068M $MnCl_2$

15% (w/v)glycerol

pH5.9

TfBII

0.01M MOPS

0.01M RbCl

0.075M $CaCl_2$

15% (w/v) glycerol

pH6.8

DTT (dithiothreitol)

2.25M in 40mM KoAc pH6.0

Solution I (GTE:glucose-Tris-EDTA)

0.05M glucose

0.024M Tris.Cl pH8.0

0.01M EDTA pH8.0

Solution II (alkaline-SDS)

0.2M NaOH (made fresh)

1% (w/v) SDS (sodium dodecyl sulphate)

This solution must be made up just prior to use.

Solution III (potassium acetate solution)

Dissolve 207g $KAc.3H_2O$ or 133g KAc in 300ml SDW.

Make up to 443ml and add 57ml acetic acid.

This solution is 3M for potassium and 5M for acetate.

Southern Blot hybridisation and prehybridisation buffer

25mM $NaPO_4$ pH7.4

5xSSC

0.5% (w/v) milk powder

50μg/ml yeast tRNA

50% (v/v) formamide

1% (w/v) SDS

Colony Screening hybridisation buffer

0.5% (w/v) milk powder

10mM Tris.Cl pH7.5

6xSSC

1mM EDTA pH8.0

1% (w/v) SDS

0.1mg/ml yeast tRNA or sonicated salmon sperm DNA

Lysis buffer for Genomic DNA Prepration

1M NaCl

50mM Tris.Cl pH8.0

1mM EDTA pH 8.0

0.5% SDS

30% acrylamide for Sequencing Gels

Make up acrylamide/N-N methylene bisacrylamide in the ratio 29:1(30% w/v).

Filter sterilise, wrap bottle in foil and store at 4°C.

Bacterial Culture Reagents

LB broth

1% (w/v) tryptone

0.5% (w/v) yeast extract

0.17M NaCl

pH7.0

2xYT broth (yeast/tryptone)

1.6% (w/v) tryptone

0.8% (w/v) yeast extract

0.085M NaCl

pH7.0

Psi broth

2% (w/v) tryptone

0.5% (w/v) yeast extract

0.01M NaCl

0.02M $MgSO_4$

0.005M KCl

'Phage Adsorption buffer

10mM $CaCl_2$

10mM $MgCl_2$

SM buffer (sodium/magnesium)

10mM Tris.Cl pH7.5

10mM $MgCl_2$

50mM NaCl

0.01% gelatin

IPTG stock solution

24mg/ml isopropylthio-β-D-galactoside in SDW.

Filter sterilise and store at $-20^{o}C$.

X-gal stock solution

40mg/nl 5-bromo-4-chloro-3-indolyl-β-D galactoside in DMF (dimethyl formamide).

Wrap container in foil and store at $-20^{o}C$.

Ampicillin stock solution

25mg/ml ampicillin in SDW.

Filter sterilise and store at $4^{o}C$.

Appendix 2.

A Sequencing Primers

Primer Name	Primer Sequence 5' 3'	Clone Sequenced
392	AACAGCTATGACCATGA	all ovine clones
393	GTAAAACGACGGCCAGT	
5801	CCTGAGTGTCCTTGATGC	
5802	AAGGAGACAGTGATGGGC	SCD1A25
B200	GTCTACGGTTCAGGTTAA	
B201	AGTGAAGCACAGCAGTCT	
G5784	GGATTTCGTGAGCATCAAG	
6088	CTGAAGCCTGGCTTTCCAGT	
6086	TGACTCCTAGGCTGACAAC	SCD1B-42
B205	TTGGTATCTCCGAGTAAC	
B204	TAGAATCCTGAGACATGG	
6087	TCCTGAAGCCCTGGTCC	
6086	AS ABOVE	
A410	CTTTAGGAGGACTGGATG	SCD1B-52
A411	GAGGAGTCTCTCAAGGAT	
G3539	GCATTATGGTTTTGGAGACG	
G3540	GCTTCAGGAAAATGGCAG	
G3537	GTGTTTTCATTACTCAGTACC	
G3538	GTTAAAATGCTGGTCATCTC	
G3774	GCCTGAGTTGTCGAGTG	SCD1T10
G3775	GGAGACATGGTCAGGGATC	
G4253	CAAATTAGTGTTCTGATAAGC	
G4252	TACCAGGCTGCTCCTCCTGCTCA	
M13for	GTAAAACGACGGCCAGT	human CD1C clone
M13rev	TTCACACAGGAAACAG	5'ends PCR clones
		gut CD1 clone

B PCR Primers

Primer Name	Primer Sequence (5' → 3')	PCR Reaction
B199	AA A GA C *** G GAAGTCA A AGAG GG G CT T	5'ends PCR gut CD1 PCR
5801	see seq. primer list	A25 α3/TM/CYT
5802	see seq. primer list	PCR
B202(GSP2)	***GCTATGTCCTGGATCACAAAGG	SCD1A25 5' end
C334(GSP1)	***CAAGAAGCTTTCTACGGCCT	PCR
B203(GSP2)	***CATCCAGTCCTCCTAAAGCT	SCD1B-42 5' end
C333(GSP1)	***GAATGATTCTGGATCCTCAA	PCR
G5933	GAGCCTGTCTATGATAGCCAAG	
G5934	ACTGAGTAATGAAAACAC	
G5936	TGAAGCCTGGCTGTCCAGT	SCD1T10 α3 PCR
G5935	CCAGTACAGGATGATATCC	
G7206	CCAAATGGAAAAGGAGAC	gut CD1 PCR
B186	GACTCGAGTCGACATCGA(Tx17)	anchored PCR
B185	GACTCGAGTCGACATCG	for 5'ends
M13for	see seq. primer list	human CD1C probe
M13rev	see seq. primer list	PCR
ATPase 1	GCTGACTTGGTCATCTGC	ATPase control
ATPase 2	CAGGTAGGTTTGAGGGGATAC	PCR
***	ATCGATGAATTC (ClaI/EcoRI enzyme adaptor seq.)	

Appendix 3: Genbank Accession Numbers

The accession numbers for human CD1 (HCD1), mouse CD1 (MCD1), rabbit CD1 (DomRab/CtRab), sheep (S) class I, and sheep (S) class II nucleotide sequences are given below. The HCD1E sequence and both rabbit sequences originated from genomic DNA whereas all the other sequences were from cDNA clones.

Molecule	Accession Number	Reference
HCD1A	M28825	Aruffo and Seed, 1989
HCD1B	M28826	Aruffo and Seed, 1989
HCD1C	M28827	Aruffo and Seed, 1989
HCD1D	J04142	Balk et al, 1989
HCD1E	X14975	Calabi et al, 1989a
MCD1.1	M63695	Balk et al, 1991
MCD1.2	M63697	Balk et al, 1991
DomRabCD1	M26248	Calabi et al, 1989b
CtRabCD1	M26249	Calabi et al, 1989b
S class I-1	M34676	Grossberger et al, 1990
S class I-2	M34675	Grossberger et al, 1990
S class I-3	M34674	Grossberger et al, 1990
S class I-4	M34673	Grossberger et al, 1990
S class II-1	M73984	none published
S class II-2	L08792	none published