University of Alberta

C-type lectin-like immunoreceptor genes of the duck *Anas platyrhynchos*

by

William George Branton

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the
requirements for the degree of *Master of Science*

Department of *Biological Sciences*

Edmonton, Alberta
Spring 2004

Your file Votre référence
ISBN: 0-612-96453-1
Our file Notre référence
ISBN: 0-612-96453-1

Abstract:

Eight unique complementary DNAs encoding C-type lectins were identified through sequencing of randomly selected clones from the duck. The first group of three clones showed homology to immunoreceptors expressed on mammalian antigen presenting cells. One member of this group had an immunoreceptor tyrosine-based inhibitory motif and one exhibited features typical of an activating receptor. Southern blot analysis indicated that each was encoded by one or two genes. These genes were differentially expressed in tissues, with the highest expression in the spleen. Transcripts for both the inhibitory and activating receptors were expressed in leukocyte populations rich in antigen presenting cells. The second group of five clones resembled mammalian leukocyte activation antigens. Southern blot analysis revealed a small family of related genes that were polymorphic between individuals. These lectin transcripts were present in all tissues tested and in various leukocyte populations.

Acknowledgements:

I would like to thank all of the people who helped me arrive at this point. First I would like to thank my supervisor, Dr. Kathy Magor for giving me the opportunity to study in her lab and pursue my interest in lectin immunoreceptors. Thank you to Kathy, to my committee members Dr. Mike Belosevic and Dr. Kevin Kane, to Dr. Brad Magor and to all members of the Magor labs, past and present for their advice and support throughout this work.

I also want to thank my friends and family for being so supportive. In particular, I would like to thank Greg Plummer for constantly helping me keep things in perspective, my grandmother, Stella Branton, and aunt Mary Zuk who's support has gotten me through both my undergraduate and graduate years. Last but not least, I want to thank my parents Brian and Karen Branton for believing in me and being such great parents to such a pain.

Table of Contents:

List of Tables:

List of Figures:

List of symbols and abbreviations:

AICL – activation induced C-type lectin

B-Lec – B locus lectin

CRD - Carbohydrate recognition domain

CLECSF – C-type lectin superfamily

CLR – C-type lectin related

ConA – concanavalin A

CMV – Cytomegalo Virus

CTLD – C-type lectin-like domain

DC – dendritic cell

DCAR – dendritic cell immunoactivating receptor

DCIR – dendritic cell immunoreceptor

DC-SIGN – dendritic cell-specific intercellular adhesion molecule 3 grabbing nonintegrin

DC-SIGNR - dendritic cell-specific intercellular adhesion molecule 3 grabbing nonintegrin related receptor

Dectin-1 – dendritic cell-associated lectin 1

Dectin-2 - dendritic cell-associated lectin 2

DLEC – dendritic cell lectin

EDTA - ethylenediaminetetraacetic acid

EGTA – ethylene glycol-bis-(b-amino-ethyl ether) N,N,N′,N′-tetra-acetic acid

ERK – extracellular regulated kinase

EST – Expressed sequence tag

FcγRIIB – Fc gamma receptor II β

GM-CSF - Granulocyte macrophage colony stimulating factor

HBV - Hepatitis B Virus

HCV – Hepatitic C Virus

HIV – Human Immunodeficiency Virus

ICAM – intercellular adhesion molecule

IFN – interferon

IL – interleukin

ITAM – Immunoreceptor tyrosine-based activation motif

ITIM – Immunoreceptor tyrosine-based inhibitory motif

LLT1 – lectin-like transcript 1

LPS – lipopolysaccahride

L-SIGN – liver/ lymph node-specific intercellular adhesion molecule 3 grabbing nonintegrin

MIIC – MHC class II compartments

MBP-A - Mannose binding protein A

MCL – macrophage C-type lectin

MGL – macrophage galactose/ N-acetylgalactosime-specific lectin

MHC – major histocompatibility complex

Mincle – macrophage inducible C-type lectin

MMR – macrophage mannose receptor

MyD88 - Myeloid differentiation primary response gene

NK – Natural Killer

NKC - Natural Killer complex

NKD – Natural Killer domains

pDCs – plasmacytoid DCs

PLA – phospholipase A

PLC – phospholipase C

PMA - phorbol 12-myristate 13-acetate

PRR – pattern recognition receptor

RT-PCR – reverse transcription polymerase chain reaction

SHP-1 – Src homology 2 domain-containing tyrosine phosphatase 1

SHP-2 - Src homology 2 domain-containing tyrosine phosphatase 2

SHIP - Src homology 2-containing-inositol 5'-phosphatase

Syk – spleen tyrosine kinase

TCR – T cell receptor

TAP – transporter associated with antigen processing

TNF – tumor necrosis factor

TLR – Toll-like receptor

Y-Lec – Rfp-Y locus lectin

1 Introduction

1.1 C-Type Lectins

C-type lectins are so named because they bind carbohydrates in a calcium dependent manner. These lectins are characterized by the presence of a common carbohydrate recognition domain (CRD) or C-type lectin-like domain (CTLD). A large superfamily of proteins feature CTLDs, though not all of them function as classical calcium dependent sugar binding proteins.

C-type lectins are characterized by CRDs which are generally between 115 and 130 amino acids in length that feature 14 invariant and 18 highly conserved amino acid residues (Drickamer, 1988) (Drickamer, 1993) with a common fold. CRDs are compact globular structures with common secondary structural features. The conserved secondary structural features include two alpha helices and two beta sheets composed of five major and several minor beta strands. The C-type lectin CRD is exemplified by the structure of the CRD of rat mannose binding protein A (MBP-A) (Weis *et al.*, 1992) (Figure 1). Most of the variation between different CTLDs lies in the various loop structures; particularly loops three and four which are either known or presumed to be involved in ligand binding in various lectins (Weis *et al.*, 1998). The structure is stabilized in part by intrachain disulphide bonds between two pairs of conserved cysteines in short form lectins and three pairs of cysteines in long form lectins. Another conserved feature of CRDs is the so-called "WIGL" motif, although only the glycine residue shows absolute conservation within the family (Drickamer, 1993). The WIGL motif forms β-strand 2 according to the numbering used for rat MBP-A, which is the linkage point between the two hydrophobic cores of the protein and is indispensable to the proper folding of the domain.

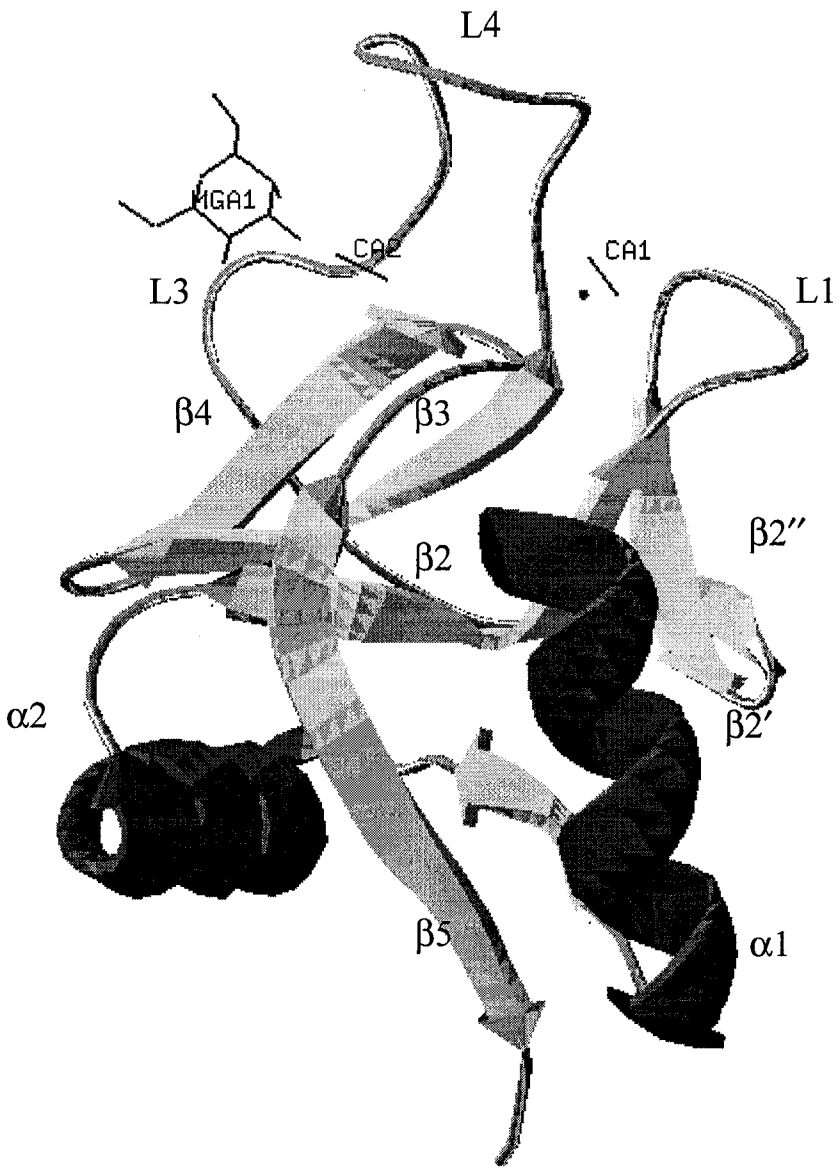


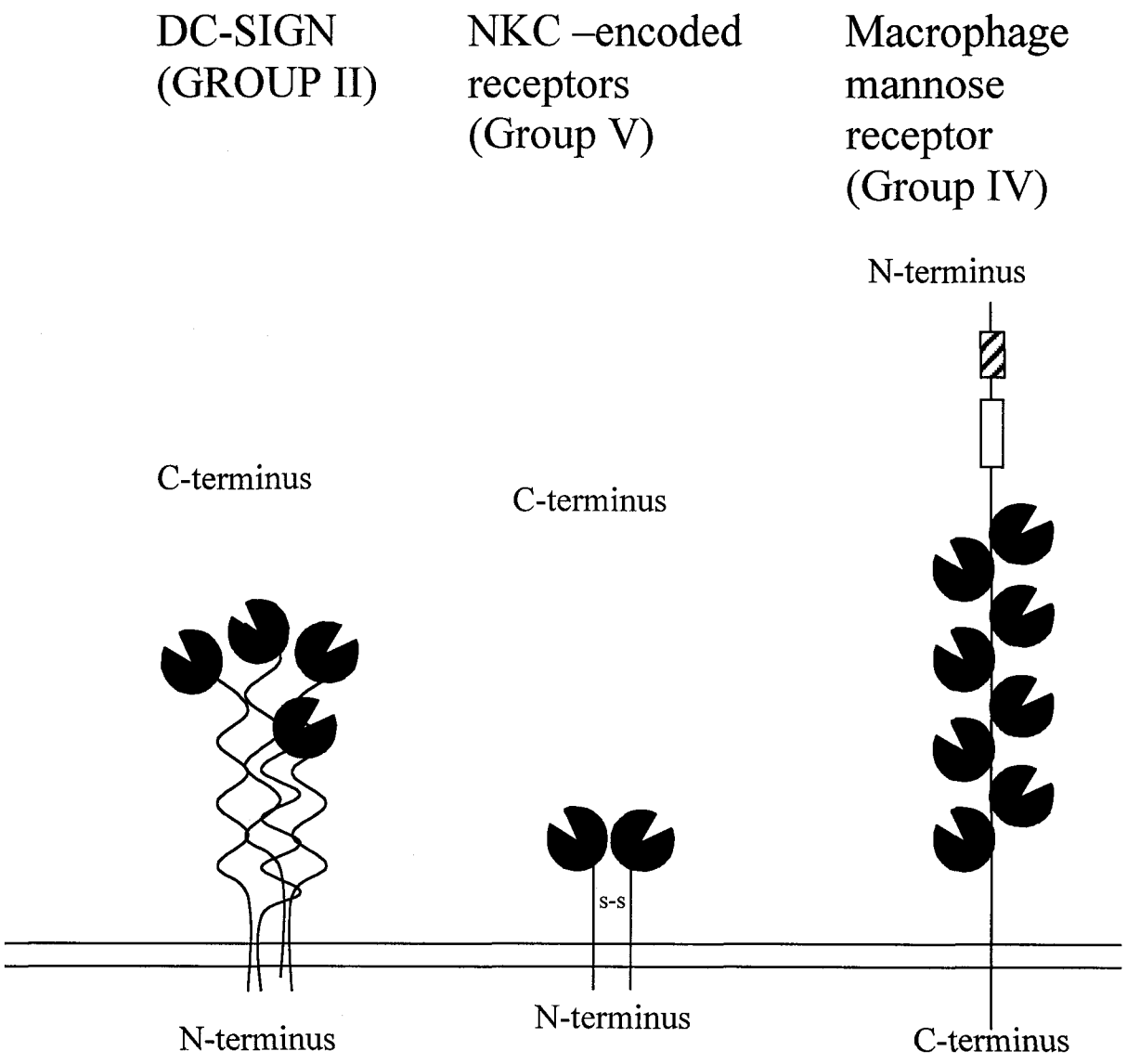
Figure 1: Crystal structure of the rat serum mannose binding protein A. CA1 represents calcium-binding site one. CA2 represents calcium-binding site two. MGA1 represents bound mannose. α1 represents alpha helix 1. α2 represents alpha helix 2. β1-5 represent the conserved beta strands. L1, L3 and L4 represent the important loop structures visible in this diagram. Image generated using the SwissPdb viewer based on pdb file associated with the Swiss-Prot entry for MBP-A.

The ability of C-type lectins to bind carbohydrates in a calcium dependent manner is attributable to nine amino acids within the CRD. These nine conserved residues form two calcium-binding sites in classical lectins, although it is only the second site that is absolutely required for calcium dependent carbohydrate binding. A tri-amino acid motif within site 2 is involved in both Ca^{2+} ligation and in determining the carbohydrate specificity of the domain. An EPN motif at this site is associated with specificity for mannose and related carbohydrates whereas a QPD motif confers specificity for galactose type sugars (Drickamer, 1992). Variants on this motif, such as EPS have been identified. The effects of such variants on binding specificity are largely unknown. The EPS motif, for example, is found in a galactose specific lectin in tunicates (Suzuki *et al.*, 1990), but has unknown specificity in mammalian lectins.

The C-type lectin superfamily can be divided into seven groups. The Group I lectins are proteoglycans and includes proteins such as aggrecan and neurocan. Group II C-type lectins are generally endocytosis receptors with a type II transmembrane orientation and a single C-terminal CRD (Drickamer and Taylor, 1993). These receptors are usually expressed on the cell surface as multimers. Group III lectins are involved in innate mechanisms of humoral defense and include the mannose-binding proteins capable of activating complement (Drickamer and Taylor, 1993). The adhesion molecules referred to as selectins comprise Group IV (Drickamer and Taylor, 1993). Group V lectins are also type II transmembrane receptors, which are usually expressed on the cell surface as dimers. Many of the Group V lectins are receptors which are encoded in the Natural Killer Complex (NKC) (Yokoyama *et al.*, 1991) (Renedo *et al.*, 1997) (Brown *et al.*, 1997) (Yokoyama and Plougastel, 2003) and are expressed on Natural Killer cells

(NK cells). These lectins feature the CRD fold but have divergent amino acid sequences that abolish the calcium binding sites characteristic of other lectins (Drickamer, 1993), thus, they are not able to bind carbohydrates in the same manner as classical lectins and many do not bind carbohydrates at all. The lectin-like domains of the Group V lectins have also been referred to as Natural Killer domains (NKD) (Weis *et al.*, 1998) due to their unique deviations from other members in the C-type lectin superfamily. Group VI lectins are receptors bearing multiple carbohydrate recognition domains on a single polypeptide with a type I transmembrane orientation (Drickamer and Taylor, 1993). Group VII is composed of secreted proteins, such as pancreatic stone protein, which consist of CRD domains independent of any other protein module. The physiological role of the Group VII lectins is unknown (Drickamer and Taylor, 1993). Examples of the structure and multimerization state of various immune type lectins are shown (Figure 2).

The group designations described above are based on protein structure, however the distinctions are also apparent in the structure of the genes encoding each group of proteins. The CRD encoding exons in each group are separated from other domains by an intron, however the intron exon organization with the CRD encoding portion of each group is unique. Groups I, II, V, and VI feature three CRD encoding exons separated by introns at positions that are conserved within a group but differ between them. On the other hand, Group VII lectin CRDs are encoded in three exons and the remaining groups feature CRDs encoded by a single exon (Drickamer, 1993).



Cysteine rich domain – also involved in carbohydrate binding

Fibronectin type II domain

Figure 2: Examples of domain organization and multimerization state of various C-type lectin receptors.

Lectins belonging to several of these groups play a wide variety of roles in the immune response. Group III lectins, for example are important innate components of humoral immunity through recognition of foreign glycoproteins and initiation of the complement cascade. Lectins within Groups I, II and V are involved in cell-cell interactions and various pathogen recognition events important for the initiation, maintenance and regulation of the immune response. The specific roles of individual lectins belonging to these three groups in pathogen recognition, cell-cell interactions and cellular signaling will be the focus of this discussion.

1.2 Immunoreceptor Tyrosine-based signaling motifs

There are two motifs shared by receptors with vastly different structures that are involved in immunoreceptor mediated signaling events, the immunoreceptor tyrosine-based activation motifs (ITAMs) and immunoreceptor tyrosine-based inhibitory motifs (ITIMs). ITAMs are found in the cytoplasmic domains of multiple receptors or adaptor proteins within receptor complexes. Upon receptor engagement, whether the ITAM motif is present on the receptor molecule itself or on an associated adaptor protein, the ITAM become phosphorylated at tyrosine residues. This phosphorylation mediates the recruitment of syk kinase and the initiation of a tyrosine kinase signal cascade resulting in the activation of gene expression and cellular effector functions. On the other hand, ITIMs are usually present on the same polypeptide as the ligand-binding portion of inhibitory receptors. ITIMs are also tyrosine phosphorylated upon receptor engagement, which typically results in SHP-1 phosphatase recruitment and activation, though SHP-2 and SHIP may also be recruited by some ITIM bearing receptors. This ultimately leads to the dephosphorylation of intracellular messenger molecules involved in tyrosine-

kinase driven activation cascades to prevent cellular activation. Often the outcomes of interactions in the immune system, such as NK cell interactions with potential targets, are determined by the balance between activating and inhibitory signaling.

1.3 Dendritic cell lectins

Dendritic cells (DCs) form a pivotal interface between the innate and adaptive immune system. Dendritic cells have two functional states: immature and mature. Each state is associated with specific functions and specific patterns of receptor expression. Receptors belonging to the C-type lectin superfamily have important roles in many aspects of dendritic cell biology in both functional states.

Immature DCs are sentinels. They are potent phagocytes present in most tissues of the body. In a healthy individual, together with macrophages, they act to clear normal debris from the body. DCs and macrophages also form a vital part of the innate barrier against infection by engulfing and destroying foreign particles that enter the body. The phagocytic functions of dendritic cells are mediated by receptors that fall into three categories: those receptors that do not distinguish between self and non-self, those that specifically recognize self-molecules and those that recognize non-self molecules. The endocytosis receptors belonging to the first category recognize a wide range of endogenous and foreign ligands and induce their uptake and targeting to early endosomes. Scavenger receptors fall into the second category. They recognize self-targets in order to direct the clearance of normal cellular debris. In the third category are the pattern recognition receptors (PRRs), which recognize molecular patterns unique to foreign particles and trigger not only their uptake and destruction, but direct them into the MHC class II processing pathway for presentation as well. Dendritic cells express

members of the C-type lectin superfamily that belong to each of these categories. In addition to their role as phagocytes, DCs are also capable of regulating the function of other effectors of innate immunity either through direct cell-to-cell contact or by the production of cytokines. Many of these functions are mediated by specific sets of lectins.

When signals of foreign invasion or damaged or altered self are detected DCs undergo a process called maturation and become potent activators of naïve T cells. DC maturation can be induced by LPS (De Smedt *et al.*, 1996), CD40-CD40-ligand interactions (Caux *et al.*, 1994), type-1 interferons (Luft *et al.*, 1998), IL-15 (Saikh *et al.*, 2001) and TNF-α. The maturation process involves a several fold reduction in endocytosis, increased surface expression of MHC class II molecules, up-regulation of T cell co-stimulatory molecules such as B7, CD83 and CD86, trafficking from the site of infection to the T-zone of draining lymph nodes and contacting naïve T cells for antigen presentation (Lane and Brocker, 1999). The latter two events are also mediated by C-type lectins. Mature DCs direct the character of subsequent adaptive immune responses through a combination of antigen presentation in the context of MHC molecules, the expression of co-stimulatory molecules and the secretion of cytokines and chemokines. DCs are also capable of interacting with effectors of the innate arm of the immune system, acting to activate Natural Killer cells and secreting cytokines to promote innate mechanisms of pathogen clearance at the site of infection.

1.3.1 Multi-CRD antigen uptake receptors

Various lectin-like receptors have been shown to have roles in several aspects of dendritic cell function. Antigen uptake and intracellular trafficking for degradation and presentation on MHC class II molecules are processes mediated by lectins, including two

members of the Group VI lectins. The Group VI lectins are type I transmembrane proteins with an N-terminal fibronectin type II repeat, a cysteine-rich region and multiple extracellular C-type lectin CRDs on a single polypeptide. This is in contrast to the other lectin groups, which achieve clustering of CRDs through oligomerization of single CRD-bearing polypeptides to form a functional receptor.

One of the Group VI lectins involved in antigen uptake by DCs is the macrophage mannose receptor (MMR). The mannose receptor bears eight extracellular carbohydrate recognition domains and all of the features of Group VI lectins described above. This receptor has a broad specificity for mannose, N-acetyl-glucosamine and other sugars. It is a classical C-type lectin in that its function is calcium dependent.

The MMR is an antigen uptake receptor. The role of the mannose receptor in antigen uptake by DCs was demonstrated through blockading this receptor with specific monoclonal antibodies, competition with mannan and addition of EDTA to deplete Ca^{2+} (Sallusto *et al.*, 1995). Each of these treatments effectively reduced the capacity of DCs to take up antigen, indicating that MMR is a major antigen uptake receptor and confirming that MMR dependent antigen uptake is mediated by calcium dependent carbohydrate binding. The efficiency of the MMR in mediating endocytosis is due, in part, to its ability to release antigen at low pH and be recycled to the surface of the cell. Thus, the receptor can capture antigen, release it into the early endosomes for degradation and be recycled to the cell surface to take up more antigen without requiring *de novo* synthesis of additional receptor (Frison *et al.*, 2003).

The second Group VI lectin expressed by DCs is DEC-205. DEC-205 is a receptor with a similar overall structure to the MMR, but has ten rather than eight

extracellular CRDs. Like the MMR, DEC-205 has a role in antigen uptake and processing. This was demonstrated using anti-DEC-205 antibodies. These antibodies specifically localized to clathrin-coated pits on the DC membrane where they were readily taken up and ultimately presented to T cells (Jiang *et al.*, 1995). This localization to coated pits is due to a conserved tyrosine-based sequence GFSSVRY that is found in the cytoplasmic domain of both the MMR and DEC-205. The ability to induce coated-pit mediated internalization is another characteristic that makes both of these molecules good antigen uptake receptors.

However, in contrast to MMR, DEC-205 targets antigens directly to the MHC class II compartments (MIIC) where antigen processing and loading onto MHC class II occurs. Thus, antigens taken up by this pathway are more readily presented to T cells (Mahnke *et al.*, 2000). This deeper targeting of DEC-205 to late endosomes/ lysosomes/ MIIC rather than early endosomes is associated with a tri-acidic motif, specifically EDE, in the cytoplasmic domain. The requirement for this tri-acidic motif is absolute as mutation of this motif to alanines abolishes the specific targeting of DEC-205 to lysosomes and reduces antigen presentation efficiency as compared to that mediated by the wildtype receptor (Mahnke *et al.*, 2000).

Typically, antigens imported from the extracellular environment are presented on MHC class II molecules, however cross-presentation of such antigens on MHC class I can occur. It has been shown that proteins taken up by lymph node resident DCs via DEC-205 can be presented on MHC class I molecules. This loading on MHC class I is accomplished in a transporter associated with antigen processing (TAP) dependent manner (Bonifaz *et al.*, 2002). The ability to present antigen on both MHC class II and

MHC class I molecules indicates that DEC-205 mediated antigen uptake can be important for the regulation of both the humoral and cellular arms of immunity.

In addition to being expressed by DCs, DEC-205 is also expressed on cortical thymic epithelial cells. Among other functions, these cells act to clear the thymus of debris from apoptotic thymocytes that have failed either positive of negative selection. This is mediated in part by DEC-205, which can bind to apoptotic thymocytes and induce their clearance (Small and Kraal, 2003).

DEC-205 mediated antigen uptake alone is not sufficient to induce DC maturation. In fact, DEC-205 mediated antigen uptake and presentation in the absence of maturation stimuli can induce tolerance (Bonifaz *et al.*, 2002). The character of the response to the endocytosed antigen is determined by the context in which these events take place. This fact coupled with the fact that DEC-205 may not distinguish between self and non-self antigens make this molecule an attractive target for therapeutic manipulation of the immune response. Ideally, DEC-205 targeted antigens could be used to induce immunity, establish tolerance or even break tolerance depending on the presence or absence of maturation stimuli delivered with the target antigen and the inoculation site.

1.3.2 DC-SIGN

DC-specific intercellular adhesion molecule 3 (ICAM-3) grabbing nonintegrin (DC-SIGN) is a Group II C-type lectin expressed on the surface of DCs. This protein is 404 amino acids in length with a type II transmembrane orientation (Geijtenbeek *et al.*, 2000c). The cytoplasmic domain is 39 amino acids long and features a dileucine motif associated with receptor internalization as well as a tri-acid motif, EEE similar to the

EDE motif of DEC-205, associated with targeting to lysosomes (Engering *et al.*, 2002a). A long neck region separates the CRD of DC-SIGN from its transmembrane region. The neck features seven imperfect repeats of 23 amino acids that facilitate tetramerization of individual subunits through the formation of alpha-helical coiled-coils (Geijtenbeek *et al.*, 2000c).

The gene encoding DC-SIGN is located on chromosome 19 linked to the related DC-SIGNR/ L-SIGN gene and the gene for the lectin-like Fc receptor CD23 (Soilleux *et al.*, 2000). An extensive repertoire of mRNAs is generated from the DC-SIGN gene by alternate splicing which result in several variant polypeptides in both soluble and membrane bound forms. These variants may form heterooligomers such that specificity altering insertions or deletions in the CRD (Mummidi *et al.*, 2001) will generate functional receptors with individual CRDs that recognize different structures. In the case of the murine homologs of DC-SIGN, there are in fact variants that lack the entire CRD (Parent *et al.*, 2002). Theoretically such truncated polypeptides could be incorporated into heterotetramers with prototypical DC-SIGN monomers and generate receptors that recognize ligands with saccharides spaced at wider intervals.

The expression of DC-SIGN is, as the name implies, restricted to dendritic cells. It mediates multiple functions on these cells including binding to naïve T cells (Geijtenbeek *et al.*, 2000c), binding to vascular endothelium (Geijtenbeek *et al.*, 2000a) and antigen uptake (Engering *et al.*, 2002a). DC-SIGN functions as a classical lectin in that it binds its ligands through carbohydrate residues and its ligand binding capacity is calcium dependent. Known ligands for DC-SIGN include intercellular adhesion molecules 2 and 3 (ICAM-2, ICAM-3).

DC-SIGN is involved in the initial interactions between DCs and naïve T cells. This lectin mediates high affinity interactions with naïve T cells, which preferentially express ICAM-3 compared to ICAM-1. DC-SIGN binding to ICAM-3 is one of the key factors that contribute to the high efficiency with which DCs are able to interact with naïve T cells. Binding of DC-SIGN to ICAM-3 can be specifically inhibited by antibodies that block either molecule and DC-SIGN interactions with ICAM-3 can be competed with mannan (Geijtenbeek *et al.*, 2000c). Thus, the binding of DC-SIGN to ICAM-3 is likely to be mediated by mannose residues on the ICAM-3 molecule. DC-SIGN mediated adhesion events are transient (Geijtenbeek *et al.*, 2000c), which ensures that interactions with T cells that are not specific for the antigen being presented can be abolished in order to allow productive interactions to have an opportunity to occur. This also supports the idea that DC-SIGN mediates initial, non-antigen specific interactions between DC and T cells that will allow later antigen specific interactions though MHC molecules and the TCR to occur.

DC-SIGN is also important for DC trafficking events. This function is mediated by strong binding to ICAM-2 (Geijtenbeek *et al.*, 2000a), which is abundantly expressed on vascular and lymphatic endothelium. As with DC-SIGN interactions with ICAM-3, these interactions can be specifically blocked with antibodies against DC-SIGN, mannan or EGTA, thus indicating that DC-SIGN binding to ICAM-2 is mediated by the carbohydrate recognition domain of DC-SIGN and that this is a Ca^{2+} dependent interaction with carbohydrates on ICAM-2. DC-SIGN is capable of mediating rolling interactions with vascular endothelium through interactions with ICAM-2. This

interaction is specific as DC-SIGN does not bind ICAM-1 and fails to bind to ICAM-3 under shear flow (Geijtenbeek *et al.*, 2000a).

DC-SIGN also functions as an antigen uptake receptor upon binding soluble antigens (Engering *et al.*, 2002a). This function is mediated by a dileucine motif in the cytoplasmic domain that allows receptor internalization upon ligand binding and a cytoplasmic, tri-acidic EEE motif, similar to the EDE motif of DEC-205, which allows DC-SIGN and bound antigen to be targeted to late endosomes/ MIIC. Upon maturation, which is accompanied by a general reduction in phagocytic activity, DC-SIGN is still internalized but is targeted to subsurface early-endosomes rather than the perinuclear late-endosomes/ MIIC (Engering *et al.*, 2002a).

Many lectins recognize the terminal sugars in complex polysaccharides. This is not the case for DC-SIGN and the closely related L-SIGN/ DC-SIGNR (Soilleux, *et al.*, 2000). Both of these lectins are unique in that they recognize internal oligosaccharide subunits within complex polysaccharide chains (Feinberg *et al.*, 2001). This is due, in part, to the relatively flat face of the CRD surrounding the site of carbohydrate recognition. This flat surface will non-specifically accommodate the residues surrounding the internal residues interacting with the binding site. Each CRD within a DC-SIGN tetramer will bind oligosaccharide subunits of a larger polysaccharide chain such that each CRD within a DC-SIGN tetramer binds a different polysaccharide chain at different glycosylation sites spaced relatively distantly on a single glycoprotein (Mitchell *et al.*, 2001). These two features provide specificity to receptors with relatively low selectivity for different sugars. The spacing of the carbohydrate residues being recognized is especially important for determining the specificity of these receptors. The

nature and spacing of sugar residues is quite different between eukaryotic and prokaryotic proteins, thus the spacing of CRDs within the tetramer allows recognition of self-glycosylation patterns such as those specific to ICAM-2 and ICAM-3, ligands of both DC-SIGN and DC-SIGNR, as opposed to bacterial glycosylation patterns.

Several viruses exploit the endocytic activity of DC-SIGN to mediate entry. Many viruses including HIV (Chehimi *et al.*, 2003) (Engering *et al.*, 2002b) (Geijtenbeek *et al.*, 2000b) (Nguyen and Hildreth, 2003) (Pohlmann *et al.*, 2001) (Trumpfheller *et al.*, 2003), Hepatitis C Virus (Pohlmann *et al.*, 2003) (Gardner *et al.*, 2003), Ebola virus (Alvarez *et al.*, 2002) (Simmons *et al.*, 2003), CMV (Halary *et al.*, 2002) and dengue virus (Tassaneetrithep *et al.*, 2003) use DC-SIGN to either enter and infect the DCs themselves or as a carrier for the virus which then infects cells, such as T cells, that the DC contacts. In fact, DC-SIGN was first identified as an HIV gp120 binding protein before its role in normal cellular function was determined.

The human L-SIGN and murine SIGNR1 and CIRE are DC-SIGN related molecules with similar functions and ligand specificities to DC-SIGN but distinct expression patterns. L-SIGN is expressed on liver sinusoidal endothelial cells (Bashirova *et al.*, 2001) and on endothelial cells in lymph nodes. The same is true for the murine protein mSIGNR1, one of several murine homologs of DC-SIGN (Park *et al.*, 2001), which is also expressed in marginal zone macrophages in the spleen and liver resident macrophages (Geijtenbeek *et al.*, 2002).

1.3.3 Dectin-1

There are several lectins expressed by dendritic cells, including dendritic cell associated lectin 1 (Dectin-1), that are encoded near genes for Group V lectins within the

NKC (Figure 3). Dectin-1 is a 247 amino acid protein with a type II transmembrane orientation bearing a single extracellular CRD. The 128 amino acid CRD contains all six cysteines conserved in the long-form CRD. The N-terminal cytoplasmic domain of Dectin-1 is 45 amino acids in length and contains an ITAM; a relatively unique feature, as most other activating lectins encoded in or near the NKC lack intrinsic signaling motifs and signal through adaptor molecules. The neck region separating the transmembrane domain and CRD of Dectin-1 is 53 amino acids in length. Unlike other NKC encoded receptors, the neck of Dectin-1 lacks a cysteine residue to form interchain disulphide bonds between members of a dimer.

Like many other genes encoding lectin-like molecules, the human Dectin-1 gene undergoes alternative splicing that either eliminates the neck encoding exon or disrupts the CRD encoding portion of the Dectin-1 message. Interestingly the splice variant lacking the neck domain encoding exon 3 is the most abundantly expressed form of Dectin-1 message (Hermanz-Falcon *et al.*, 2001). The neck region, involved in multimerization in many other lectins, is also assumed to provide conformational flexibility on the receptor allowing the CRD to bend relative to the transmembrane region in order to contact ligand. The absence of a neck region would bring the CRD nearer to the cell membrane and possibly result in a receptor with increased rigidity.

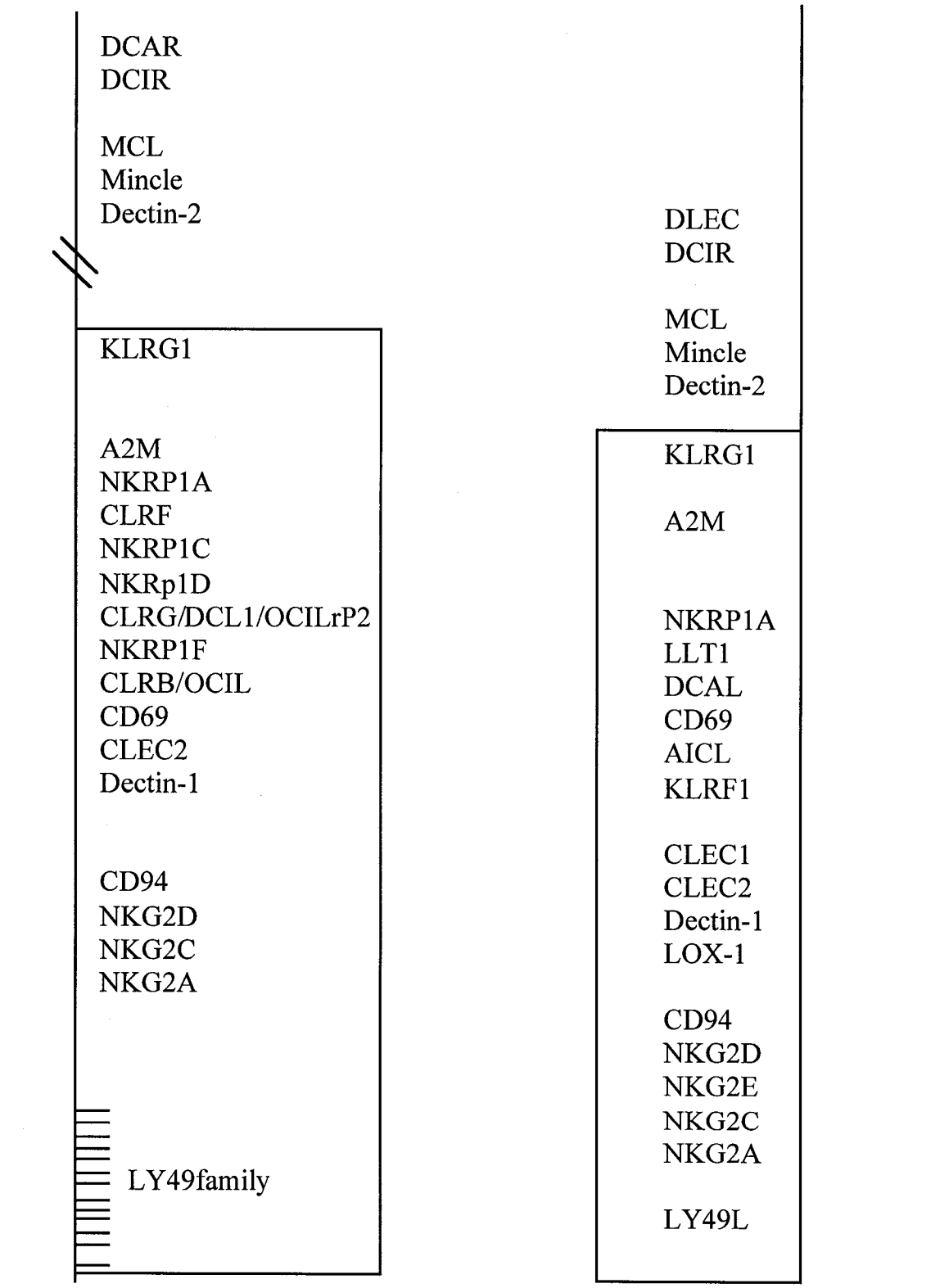


Figure 3: Schematic of the Natural Killer Complex and adjacent C-type lectin encoding genes on human chromosome 12 and mouse chromosome 6. The region classically referred to as the NKC is indicated by boxes. Dashes represent greater physical distance than depicted. Abbreviations used: DCAR: dendritic cell immunoactivating receptor, DCIR: dendritic cell immunoreceptor, MCL: macrophage C-type lectin, Mincle: macrophage inducible C-type lectin, Dectin-2: dendritic cell-associated lectin two, KLR: killer cell lectin-like receptor, A2M: alpha two microglobulin, NKR-P1: natural killer cell receptor protein 2, CLR: C-type lectin related receptor, LLT1: lectin-like transcript 1, DCAL: dendritic cell associated lectin, OCIL: osteoclast inhibitory lectin, AICL: activation induced C-type lectin, CLEC: C-type Lectin, LOX-1 low-density oxidized lipoprotein receptor, NKG2: natural killer cell receptor gene two.

Both the human and murine forms of Dectin-1 are preferentially, but not exclusively, expressed by subsets of tissue and lymphoid organ resident DCs and at a lower level on monocytes and macrophages (Hermanz-Falcon *et al.*, 2001) (Ariizumi *et al.*, 2000b). Independent experiments demonstrated that Dectin-1 expression is not restricted to dendritic cells; it is also expressed on neutrophils, a small subset of splenic T cells and is expressed on macrophages at significant levels (Taylor *et al.*, 2002) (Brown *et al.*, 2002). Expression of Dectin-1, at least on macrophages, is enhanced by IL-4 and IL-13 as well as GM-CSF and TGF-β and down regulated by IL-10 and LPS (Willment *et al.*, 2003).

Dectin-1 appears to have at least two distinct functions. Initial experiments showed carbohydrate independent binding of murine Dectin-1 to T cells, which induced enhanced proliferation *in vitro* in response to sub-optimal concentrations of anti-CD3 antibody, implying that this molecule acts in a co-stimulatory capacity (Ariizumi *et al.*, 2000b). Dectin-1 also acts as a β-glucan receptor. It was demonstrated that Dectin-1 acts as a pattern recognition receptor by binding 1,3 and 1,6 linked β glucans from fungi and plants, and mediates internalization of these ligands (Brown and Gordon, 2001). In keeping with initial findings (Ariizumi *et al.*, 2000b), Dectin-1 was not capable of binding monosaccharides or any carbohydrate structures other than β-glucans. It was also demonstrated that Dectin-1 binding to T cells and binding to β-glucans are mediated by different sites on the receptor, as β-glucans do not block binding of Dectin-1 to T cells (Brown and Gordon, 2001). Dectin-1 mediated binding to β-glucans on macrophages induces the production of TNF-α by these cells (Brown *et al.*, 2003). TNF-α is essential in anti-fungal immune responses, thus the detection of the fungal β-glucans by Dectin-1

activates a key factor involved in the innate immune response against fungal pathogens. This response is mediated by the immunoreceptor tyrosine-based activating motif (ITAM) motif in the cytoplasmic tail of Dectin-1, as recombinant proteins lacking the ITAM sequence were not able to induce TNF-α production. This effect is also dependent on Toll-like receptor 2 (TLR-2), as MyD88 and TLR-2 knockout mice lost the ability to induce TNF-α production in response to β-glucans. TLR2 and Dectin-1 co-localize at the cell surface and this is where the induction of TNF-α production occurs since inhibition of phagocytosis did not abolish induction of TNF-α (Brown *et al.*, 2003). This is in contrast to many other TLR recognition events, which occur in the phagosome. Dectin-1 and TLR-2 have a synergistic effect in inducing cytokine production. TLR2 activation in response to zymosan and Dectin-1 activation in response to zymosan are dissociable events, but neither stimulus alone is sufficient to induce significant levels of proinflammatory cytokine production (Gantner *et al.*, 2003). These receptors also induce DC maturation and specifically direct the DCs to present fungal antigens to T cells in order to initiate an immune response against the detected pathogen.

1.3.4 Dectin-2

Dendritic cell associated lectin 2 (Dectin-2) (Ariizumi *et al.*, 2000a), also known as NKCL (Fernandes *et al.*, 1999), is a DC expressed receptor encoded by a gene near the NKC. This is in contrast to the Dectin-1 gene, which is actually within the NKC proper. In fact, despite their similar nomenclature, the sequences of Dectin-1 and Dectin-2 are quite divergent. Dectin-2 is a 210 amino acid type II transmembrane protein with a short cytoplasmic domain, a single transmembrane domain a neck region and a C-terminal CRD. The 17 amino acid cytoplasmic domain of Dectin-2 lacks any intrinsic tyrosine-

based signaling motifs, nor is there a charged residue in the transmembrane domain of this receptor that would allow it to associate with adaptor proteins, thus it is unlikely that this protein alone is capable of signal transduction (Ariizumi *et al.*, 2000a). The CRD of Dectin-2 retains the residues that form Ca^{2+} binding site 2 but lack those that form site 1. Dectin-2/ NKCL features the EPN motif that confers mannose specificity and that specificity has been confirmed through *in vitro* assays (Fernandes *et al.*, 1999). As with many of the related lectin encoding genes the Dectin-2 message is alternatively spliced to generate multiple isoforms (Ariizumi *et al.*, 2000a).

The function of Dectin-2 is only poorly characterized. This receptor is proposed to have a role in radiation-induced tolerance. Radiation induced tolerance describes a phenomenon whereby lasting tolerance is induced against antigens introduced at a site of UV irradiation. This is assumed to be the result of inappropriate activation of regulatory T cells. Dectin-2 may induce this tolerance through interactions with a ligand expressed by regulatory T cells, as addition of soluble Dectin-2 acts as an antagonist preventing and even breaking UV induced tolerance (Aragane *et al.*, 2003).

1.3.5 DCIR

Many other lectins are known to be expressed by DCs that have unknown functions. One of these is the dendritic cell immunoreceptor (DCIR), which has been independently cloned by several groups (Bates *et al.*, 1999) (Huang *et al.*, 2001) (Richard *et al.*, 2002). The DCIR gene was mapped near the natural killer complex of genes in both mice and humans (Bates *et al.*, 1999) (Huang *et al.*, 2001). The gene is divided into six exons. Exon 1 encodes the cytoplasmic domain, exon two encodes the transmembrane domain, the neck region is encoded by exon three and the remaining three

exons make up the CRD encoding region. The mRNA is approximately 1300 bases in length featuring an open reading frame of 711 bases in humans and 714 bases in mice. The message also features a single *attta* motif in its 3′ untranslated region that is associated with rapid turnover of transcript. There are two known splice variants of the DCIR message that encode polypeptides lacking a TM region (Huang *et al.*, 2001) (Richard *et al.*, 2002).

DCIR is a type II transmembrane protein 237 amino acids in length in humans and 238 amino acids long in mice with a single C-terminal CRD. Within the cytoplasmic domain is an immunoreceptor tyrosine based inhibitory motif of the sequence: EITYAEV. This ITIM sequence is conserved between the murine and human homologues. There is capacity for considerable variation within the ITIM consensus of XI/LXYXXV/L (Vivier and Daeron, 1997), which makes this conservation striking. This ITIM is functional, as demonstrated by experiments using chimeric receptors bearing the cytoplasmic domain of DCIR fused with FCγRIIB. This chimeric receptor was expressed in B cells and its ability to inhibit BCR mediated activation events was tested. These experiments showed that the ITIM of DCIR is functionally capable of inhibiting cellular activation events, including tyrosine phosphorylation and Ca^{2+} mobilization. This effect was dependent on the presence of the tyrosine residue within the ITIM as mutation of this residue abolished inhibitory signaling (Kanazawa *et al.*, 2002).

The CRD of DCIR is similar to that of various macrophage associated lectins. It retains the second of two Ca^{2+} binding sites characteristic of classical C-type lectins. It is this second site that is indispensable for Ca2+-dependent carbohydrate binding; thus, DCIR is likely to function as a carbohydrate binding lectin. This is in contrast to the

natural killer cell associated lectins encoded nearby, which lack the Ca^{2+}- binding site and are involved in direct protein-protein interactions. Among the residues conserved in the binding site of classical lectins is a three amino acid motif that determines the carbohydrate specificity of the receptor. In DCIR, an EPS motif occupies this position. The specificity of this motif in mammals is unknown. Though it is more similar to the EPN motif known to confer specificity for mannosylated targets, this motif has also been found in a galactose specific lectin in tunicates (Suzuki *et al.*, 1990).

Separating the transmembrane region from the CRD is a 36 amino acid neck region. This neck region is only poorly conserved between mouse and human DCIR, however there is a conserved cysteine within this region that is likely to be involved in interchain disulphide bond formation between members of a dimer (Bates *et al.*, 1999). This mode of dimerization is common for many of the lectins encoded within the NKC.

Expression of DCIR was detectable in the spleen, lymph node, and bone marrow and was especially abundant in peripheral blood leukocytes. This expression pattern was further refined to demonstrate DCIR expression by immature DCs, CD14+ monocytes, CD19+ B lymphocytes and CD15+ blood granulocytes. However, DCIR expression was not detectable in CD3+ T cells or CD16+/ CD 56+ NK cells (Bates *et al.*, 1999). This implies that DCIR has a function specific to cells capable of antigen presentation.

The timing of DCIR expression has also been studied. In DCs, DCIR expression at the mRNA and protein level is high in immature CD14+ DCs and slightly lower in CD1a+ DCs. Maturation stimuli such as LPS, CD40/CD40L interactions and type I IFNs downregulate cell surface expression of the DCIR protein and over the course of 24 hours

after CD40/CD40L interactions DCIR mRNA expression is lost from *in vitro* generated human DCs (Bates *et al.*, 1999) (Huang *et al.*, 2001).

Another group has focused on the expression of DCIR, also known by the HUGO approved name CLECSF6, for C-type lectin superfamily member 6, in neutrophils. Resting neutrophils express low levels of both full length and truncated DCIR mRNA. Parallel to the findings in dendritic cells, upon activation by proinflammatory stimuli, neutrophils lose the surface expression of DCIR. Interestingly stimulation with cytokines such as IL-3, IL-4 and IL-13 were shown to cause accumulation of the short form of the DCIR mRNA, which lacks the transmembrane encoding exon (Richard *et al.*, 2002). This may represent a neutrophil specific way of regulating the surface expression of DCIR.

The ligand and physiological role of DCIR remains unknown, but the expression pattern of this receptor and its inhibitory capacity suggest that it plays a role during the antigen uptake phase of antigen presenting cells, or in the regulation of activation or maturation events. It is intriguing that maturation stimuli abolish DCIR expression; this implies that it inhibits a function in immature cells that is required in mature or activated cells. The similarity of the CRD region of the DCIR protein to antigen uptake receptors implies that it plays a role in surveying antigens being taken up and inhibiting activation or maturation events upon recognizing its ligand.

1.3.6 DCAR and DLEC

In mice, a direct activating counterpart to DCIR called dendritic cell immunoactivating receptor (DCAR) has been identified (Kanazawa *et al.*, 2003). The gene encoding DCAR is located 51.5kb from the DCIR gene further from the NKC. The

organization of the gene is very similar with exon 1 containing the 5′ untranslated region and encoding most of the cytoplasmic domain, exon two encodes the remainder of the cytoplasmic domain and the transmembrane domain, exon three encodes the neck and exons four, five and six encode the CRD.

DCAR features a CRD that is highly similar to that of DCIR. However, DCAR features a considerably shorter cytoplasmic domain than DCIR and lacks intrinsic signaling motifs. There is a positively charged arginine residue within the transmembrane domain. As such, the DCAR protein resembles an activating member of an activating/ inhibitory receptor pair as seen for several members of the Ly49 family of natural killer receptors. The inhibitory members of such pairs feature longer cytoplasmic domains with one or two ITIM motifs, while the activating members of the pair feature highly similar CRDs but have short cytoplasmic domains and use a single positively charged residue within their transmembrane domains to associate with ITAM bearing adaptor proteins such as DAP12 (Figure 4). Through co-transfection of 293 T cells with the DCAR cDNA and a panel of potential adaptor molecules including DAP12, DAP10, CD3ζ chain and FcRγ chain, it was demonstrated that DCAR associates with FcRγ chain to form a functional receptor (Kanazawa *et al.*, 2003). Through this association, DCAR is capable of inducing calcium mobilization and tyrosine phosphorylation of cellular proteins to mediate cellular activation.

The expression of DCAR is distinct from that of DCIR at the tissue level. DCAR expression is strongest in the lung and spleen with weaker expression in the skin and lymph node (Kanazawa *et al.*, 2003). This is in contrast to the broader expression of DCIR. This difference in the expression pattern between DCAR and DCIR supports the

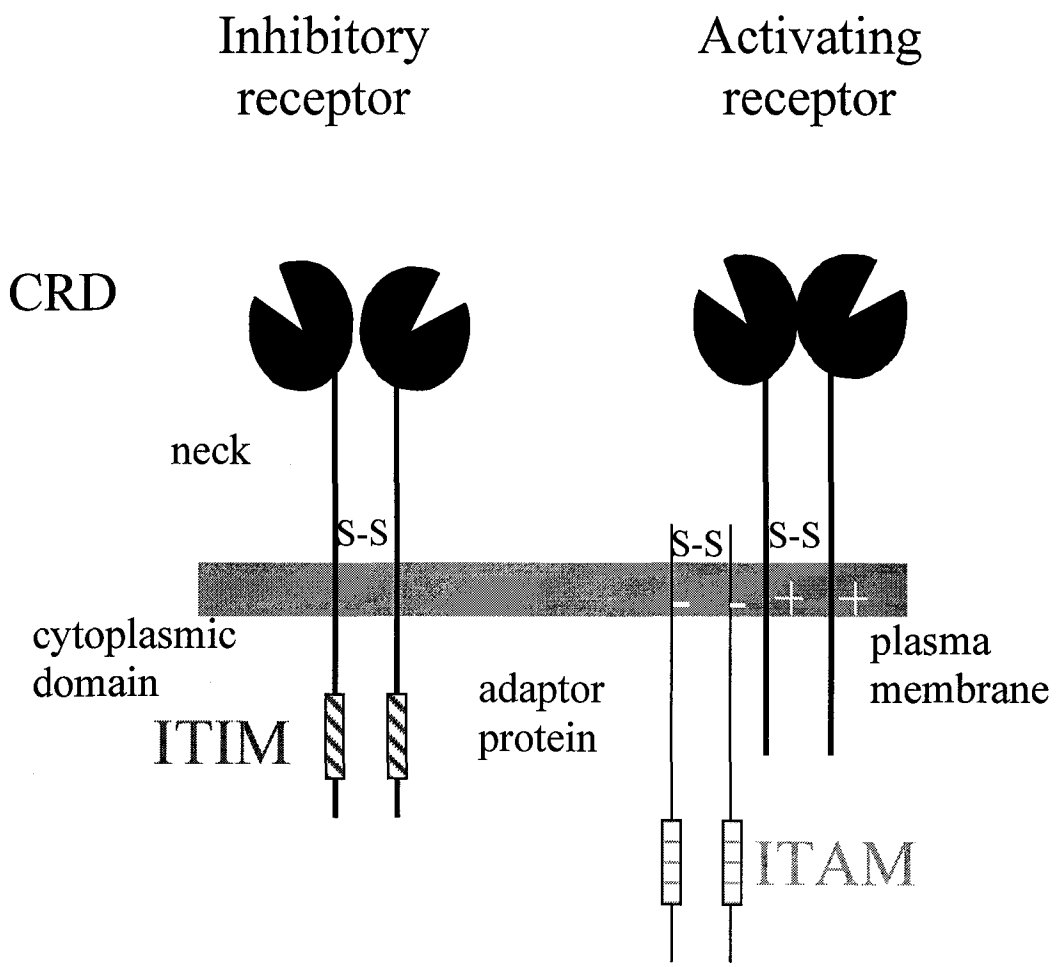


Figure 4: Schematic of activating and inhibitory lectin-like immunoreceptors. S-S represents interchain disulphide bonds. ITIM: immunoreceptor tyrosine-based inhibitory motif. ITAM: Immunoreceptor tyrosine based activation motif.

possibility that DCAR mediated activation is only appropriate in certain contexts whereas the broader expression of DCIR suggests that its function is more widely relevant. The effect of activation events on DCAR expression was not examined, but DCAR expression was detected on immature DCs. (Kanazawa *et al.*, 2003).

A direct activating counterpart for DCIR has yet to be identified in humans, however, a protein related to DCIR, but lacking an ITIM has been identified (Arce *et al.*, 2001). This protein, called dendritic cell lectin (DLEC) is similar to DCAR in many respects, however it does not carry a charged residue in its transmembrane domain and therefore cannot form an activating receptor by association with adaptor molecules in the same manner. Like DCAR in mice, the gene for DLEC is found linked to the human DCIR gene, and features the same gene organization. It also bears a reduced cytoplasmic domain when compared to DCIR and lacks intrinsic signaling motifs. The CRDs of DCIR and DLEC are 79% identical. Like DCIR, DLEC retains the residues that form Ca^{2+} binding site 2 but lack site 1. The carbohydrate specificity-determining motif in DLEC is EPN, which confers specificity for mannose.

In contrast to other related genes, the expression of DLEC mRNA was not detectable in any tissue by northern blotting implying that this is a rare transcript. RT-PCR experiments using 32 cycles showed expression in PBMCs and monocyte derived immature DCs. Thirty-five cycle experiments showed DLEC expression in macrophages and monocytes as well, but as it is undetectable at fewer cycles, this transcript is even more rare in these cells. Northern blots of RNA extracted from pure cell populations show that immature DCs are the only cells that express DLEC message at a level

detectable by this method. Expression of DLEC is lost upon maturation, as is the case with DCIR (Arce *et al.*, 2001).

1.3.7 BDCA-2

BDCA-2 is a DC associated lectin with a type II transmembrane orientation (Dzionek *et al.*, 2001). BDCA-2 is similar in structure, sequence and topology to DLEC. However, unlike the former, BDCA-2 is preferentially expressed by plasmacytoid DCs (pDCs). Monoclonal antibodies directed against BDCA-2 were readily internalized and presented to CD4+ T cells indicating that BDCA-2 is capable of internalizing bound antigens and directing them to the MHC class II pathway (Dzionek *et al.*, 2001). It was also demonstrated that binding of BDCA-2 induced intracellular calcium mobilization and protein tyrosine kinase activation (Dzionek *et al.*, 2001). In contrast to its apparent role in cellular activation, cross-linking BDCA-2 inhibits IFN α and IFN β production even when the pDCs are treated with strong stimulators of IFN production such as CpG DNA (Dzionek *et al.*, 2001). This is intriguing, BDCA-2 lacks intrinsic signaling motifs and was detected as a monomer even under non-reducing conditions (Dzionek *et al.*, 2001) so the capacity of BDCA-2 to mediate signaling implies the use of a novel signaling mechanism rather than heterodimerization with a typical signaling polypeptide.

1.3.8 DCAL-1

The DCAL-1 gene, located within the human NKC near the CD69 gene, encodes an additional C-type lectin superfamily member with a type II transmembrane orientation and preferential expression on antigen presenting cells. The DCAL-1 protein bears a CTLD that is quite divergent from classical lectins: the domain lacks both of the two

calcium binding sites, 3 of the 6 cysteines involved in intrachain disulphide bond formation, and the WIGL motif which forms the hydrophobic core of the CRD.

DCAL-1 mRNA is expressed by immature and mature DCs and CD40 stimulated B cells. At the protein level, B cells and $CD1C^{+}$ $CD11c^{bright}$, $CD123^{dim}$ myeloid DCs showed high levels of DCAL-1 protein at the cell surface, whereas plasmacytoid DCs did not. DCAL-1 binds $CD4^{+}$ $CD45RA^{+}$ T cells, as well as CD20+ peripheral blood B cells. As with Dectin-1, binding of DCAL-1 to T cells has a co-stimulatory function *in vitro*, even at low concentrations of anti-CD3 antibody. DCAL-1 co-stimulation specifically enhanced the production of IL-4, a TH2 cytokine, by these T cells (Ryan *et al.*, 2002) although the physiological relevance of this remains unknown.

1.3.9 CLEC-1

CLEC-1 is another C-type lectin with a presumed type II transmembrane orientation. This gene is located within the human NKC region linked to the scavenger receptor LOX-1, an additional related lectin CLEC-2, which shows a less restricted expression pattern, and Dectin-1 (Colonna *et al.*, 2000). This molecule has been identified in the NKC of humans but not that of mice, unlike Dectin-1 and CLEC-2, which exist in a region of otherwise conserved synteny in both organisms. CLEC-1 mRNA is preferentially expressed in mature DCs. CLEC-1 lacks the residues involved in calcium ligation, thus CLEC-1, like many NKC encoded lectins, is unlikely to function as a calcium dependent carbohydrate-binding molecule. CLEC-1 features the sequence YSST in its cytoplasmic tail. This does not match the consensus sequence for ITIM, ITAM or the tyrosine based internalization motif consensus, but the presence of a

tyrosine, multiple serines and a threonine in the cytoplasmic domain are likely to have functional importance.

CLEC-1 is not expressed at the cell surface as a monomer or homodimer. In order to confirm the predicted type II transmembrane orientation of CLEC-1, it was expressed in COS cells. It was found that despite the presence of a transmembrane domain CLEC-1 is not detectable at the surface of transfected cells, thus it either has an intracellular function or it is only expressed on the cell surface as a heterodimer with a molecule that is not expressed in COS cells (Colonna, *et al.*, 2000).

1.4 Macrophage C-type Lectins

1.4.1 MCL/ CLECSF8

In addition to the multi-CRD macrophage mannose receptor, macrophages also preferentially express lectins bearing a single CRD. One of these, simply called macrophage C-type lectin (MCL) in mice and CLECSF8 in humans is a receptor with a type II transmembrane orientation. As is the case with DCIR, the gene encoding MCL is located near to, but not within the NKC region in both mice (Balch *et al.*, 2002) and men (Arce *et al.*, 2004). In mice, the MCL message, unlike other related genes, has not been demonstrated to undergo alternate splicing events with the exception of the rare deletion of the first glycine encoding codon of exon IV in some transcripts. However, some transcripts have been identified which use a non-canonical polyadenylation signal that results in an extended 3′ UTR (Balch *et al.*, 1998). The human CLECSF8/ MCL gene on the other hand, feature a splice variant lacking a large portion of exon IV that introduces a stop codon resulting in a polypeptide lacking a CRD (Arce *et al.*, 2004). The murine MCL mRNA features an ORF of 660 bp encoding a 219 amino acid type II

transmembrane receptor with a C-terminal CRD (Balch *et al.*, 1998). The human MCL ortholog known as CLECSF8 is a 215 amino acid protein (Arce *et al.*, 2004).

The MCL CRD shows conservation of the residues that form both of the calcium co-ordination sites found in classical lectins, however the EPN motif associated with binding to mannose-type ligands in other C-type lectins is replaced by an ESN motif (Balch *et al.*, 1998). This may abolish the sugar-binding site in this receptor due to the significant difference between the structures of proline and serine, especially in terms of stereochemistry and hydrophobicity.

In humans, CLECSF8 message is strictly restricted to monocytes and macrophages and, as is the case in mice, particularly abundant in peritoneal macrophages. CLECSF8 message is upregulated by IL-6, TNF-α, IL-10 and IFNγ and downregulated by LPS. CLECSF8 expression is also lost after culture in the absence of these stimuli, implying that *in vivo* expression is maintained through environmental signals and is not constitutive (Arce *et al.*, 2004).

MCL is able to mediate endocytosis. Cells transfected with FLAG-tagged MCL readily endocytosed anti-FLAG antibodies. It remains to be determined whether MCL directs antigens to the MIIC or to early endosomes.

1.4.2 Mincle

Another lectin preferentially expressed on macrophages is the macrophage inducible C-type lectin (Mincle). Mincle is also encoded by a gene in the same region just outside the NKC as MCL, DCAR and DCIR. However, unlike the related genes encoded nearby, expression of Mincle is enhanced, rather than reduced, by LPS. In fact, Mincle expression by macrophages is known to be induced by a broad spectrum of

inflammatory stimuli. This induction is accomplished, in part, by the transcription factor NfIL-6, a transcription factor that becomes active in response to inflammatory stimuli, driving the expression of the Mincle gene (Matsumoto *et al.*, 1999).

The Mincle protein itself is structurally similar to related receptors like MCL and DLEC. It is a type II transmembrane protein with a neck region separating the transmembrane domain from a C-terminal CRD. The Mincle CRD retains the residues required for calcium coordination and thus Mincle is likely to function as a classical C-type lectin. The cytoplasmic domain of Mincle lacks tyrosine-based signaling motifs in its short cytoplasmic domain, but it does feature two potential PKC phosphorylation sites (Matsumoto *et al.*, 1999). This potential for phosphorylation may confer Mincle with signal transducing capacity; however there is no evidence that the potential sites are in fact phosphorylated, nor has the ligand for this receptor been discovered and its function is not known.

1.4.3 MGL

Macrophage galactose lectin (MGL) in mice and human macrophage lectin (HML) in humans are type II transmembrane proteins bearing a single C-terminal CRD. These receptors bear a QPD motif that confers specificity for galactose and its derivatives. It has been demonstrated that the receptor preferentially recognizes α-GalNAc over other galactose derivatives and that it is incapable of binding to mannose or related structures. This receptor is an endocytic receptor that mediates pathogen recognition and phagocytosis. Some tumors express glycoproteins with truncated polysaccharide chains known as Tn antigen. HML is also capable of recognizing Tn antigens and allows binding of tumorcidal macrophages to tumors through the

recognition of this tumor associated glycosylation pattern. As with the macrophage mannose receptor MGL is also expressed by immature DCs and acts as an antigen capture receptor directing bound antigen to the MIIC compartment for processing and loading onto MHC class II molecules. MGL expression is downregulated upon maturation, along with general phagocytic activity (Denda-Nagai *et al.*, 2002).

1.5 Activation induced C-type lectins

1.5.1 CD69

CD69 has been cloned in humans (Hamann *et al.*, 1993), mice (Ziegler *et al.*, 1994), cattle (Ahn *et al.*, 2002) and swine (Yim *et al.*, 2002), with potential homologs also identified in *Xenopus*. CD69 is a 199 amino acid type II transmembrane receptor with a 40 amino acid cytoplasmic domain and an extracellular C-type lectin carbohydrate recognition domain (Hamann, *et al.*, 1993). It is expressed as a disulphide-linked homodimer on the cell surface. Cysteine 68 in the brief neck region of CD69 is responsible for this dimerization (Sancho *et al.*, 2000). These dimers are also stabilized by hydrophobic interactions between β strand 0 and α helix 2 (Natarajan *et al.*, 2000) of each member of the dimer. The CD69 molecule is constitutively phosphorylated (Hara *et al.*, 1986) (Lanier *et al.*, 1988) at one or more serines within the cytoplasmic domain (Hamann, *et al.*, 1993).

Like many related lectins, the gene encoding CD69 is located within the natural killer complex (Ziegler *et al.*, 1994). The gene is arranged in five exons, with the exons 1 and 2 encoding the cytoplasmic and transmembrane domains and exons 3-5 encoding the CRD (Ziegler *et al.*, 1994) (Santis *et al.*, 1994). CD69 exists as a single copy gene although related genes are closely linked to it in the NKC of both mice and humans. The

CD69 gene is 15 kb in humans (Santis *et al.*, 1994) and only 7.5 kb in mice (Ziegler *et al.*, 1994). The translational start site is not in a Kozak consensus, a feature common to many lectin-encoding genes.

CD69 is expressed widely throughout the hematopoietic lineage. CD69 surface expression is induced upon activation in lymphocytes, macrophages, neutrophils, eosinophils and mast cells (Escribano *et al.*, 1997) (Sancho *et al.*, 2000), however it is induced more rapidly and expressed to a greater extent in T cells than B cells. CD69 expression is induced through a protein kinase C (PKC) dependent pathway as demonstrated by treatment with the PKC inhibitor H7. Increased intracellular Ca^{2+} enhances PKC driven CD69 expression, but is not absolutely required nor is it sufficient to induce CD69 expression in the absence of PKC signaling. CD69 is also constitutively expressed on platelets (Testi *et al.*, 1990), Langerhans cells, monocytes (De Maria *et al.*, 1994) and myeloid cell precursors.

CD69 is capable of signal transduction. Early experiments showed that cross-linking CD69 with specific monoclonal antibodies could trigger cytolytic activity of NK cells and γ/δ T cell, but not α/β T cells (Moretta *et al.*, 1991). Interestingly only mature α/β T cells showed this unresponsiveness to CD69 cross-linking, as double negative thymocytes show CD69 dependent cytotoxicity *in vitro* (Moretta *et al.*, 1991). CD69 cross-linking also enhances T cell proliferation in response to phorbol 12-myristate 13-acetate (PMA) (Cebrian *et al.*, 1988) (Nakamura *et al.*, 1989). This is mediated by upregulating the production of IL-2 and the upregulation of the IL-2 receptor as blocking IL-2R inhibits CD69 mediated proliferative responses (Cebrian *et al.*, 1988) (Testi *et al.*, 1989). Mechanistically this is due in part to CD69 induced Ca^{2+} influx supporting PKC

activation, leading to expression of the IL-2 and IFN-γ genes (Testi *et al.*, 1989), thus CD69 mediated signaling is necessary but not sufficient for induction of the IL-2 response *in vitro*.

On platelets, as with many other CD69 expressing cells cross-linking antibody against CD69 induced cellular activation events. In the case of platelets, CD69 cross-linking induced aggregation and degranulation as well as activation of the cyclooxygenase pathway (Testi *et al.*, 1990). Platelet activation induced by CD69 is mediated by phospholipase A2 activation. PLA2 cleaves phosphatidylcholine and phosphatidylethanolamine to release arachidonic acid (AA), which is processed by the cyclooxegenase pathway to generate response mediators (Testi *et al.*, 1992). These mediators then act to induce aggregation and degranulation.

In monocytes, CD69 cross-linking is also associated with Ca^{2+} influx and phospholipase A2 activation (PLA2) (De Maria *et al.*, 1994). The production of AA by CD69 activated PLA2 in monocytes causes the release of various leukotrienes and prostanoids with proinflammatory effects. CD69 cross-linking also stimulated monocytes to produce NO, and induces NO-dependent cytotoxicity of monocytes (De Maria *et al.*, 1994). CD69 expression on neutrophils has been suggested to be a mediator of activation induced apoptosis (Foerster *et al.*, 2002).

In NK cells, IL-2 treatment induces the expression of CD69 and this correlates with the lymphokine activated killer cell phenotype of killing cells previously resistant to unstimulated NK cells (Lanier *et al.*, 1988). NK cells are also activated by CD69. CD69 can induce proliferation, cytotoxicity and TNFα production in NK cells stimulated with IL-2 or IL-12. These functions are accomplished via CD69 induced Ca^{2+} influx in NK

cells (Borrego *et al.*, 1999). In NK cells CD69 mediated activation can be suppressed by ITIM signaling via CD94/NKG2A (Borrego *et al.*, 1999) (Zingoni *et al.*, 2000).

CD69 signaling has been attributed to association with G-coupled proteins (Risso *et al.*, 1991) (Sancho *et al.*, 2000), ERK activation (Zingoni *et al.*, 2000) and Syk activation (Pisegna *et al.*, 2002), which is a probable mechanism by which CD69 induces ERK activation. CD69 also induces PLCγ activation (Pisegna *et al.*, 2002). Though the exact pathway remains elusive, a possible model of CD69 mediated signaling is presented (Figure 5).

The nature of the ligand for CD69 is a matter of controversy. The amino acid sequence of CD69 suggests that it is not a typical calcium dependant lectin and may be involved in protein-protein interactions. However, biochemical analysis suggests that CD69 is in fact capable of calcium dependent binding to N-acetylglucosamine and N-acetylgalactosamine (Bezouska *et al.*, 1995) (Pavlicek *et al.*, 2003). CD69 is capable of binding calcium (Pavlicek *et al.*, 2003) despite the lack (Hamann, *et al.*, 1993) (Ziegler *et al.*, 1994) of all but one of the residues involved in calcium binding in classical C-type lectins (Drickamer, 1993) (Zelensky and Gready, 2003). This calcium binding is through a unique site involving the conserved Asp 171 and the carboxylic groups of glutamate 185 and 187 (Pavlicek *et al.*, 2003). Three distinct binding sites for N-acetylglucosamine were identified through the combination of direct binding assays with wildtype and mutant forms of the CRD and structural modeling. Two of these sites are calcium dependent while the third is not, but binds with low affinity (Pavlicek *et al.*, 2003). Unlike the binding observed in classical C-type lectins, carbohydrate binding in CD69 lacks direct association between the carbohydrate residue and the ligated calcium; rather

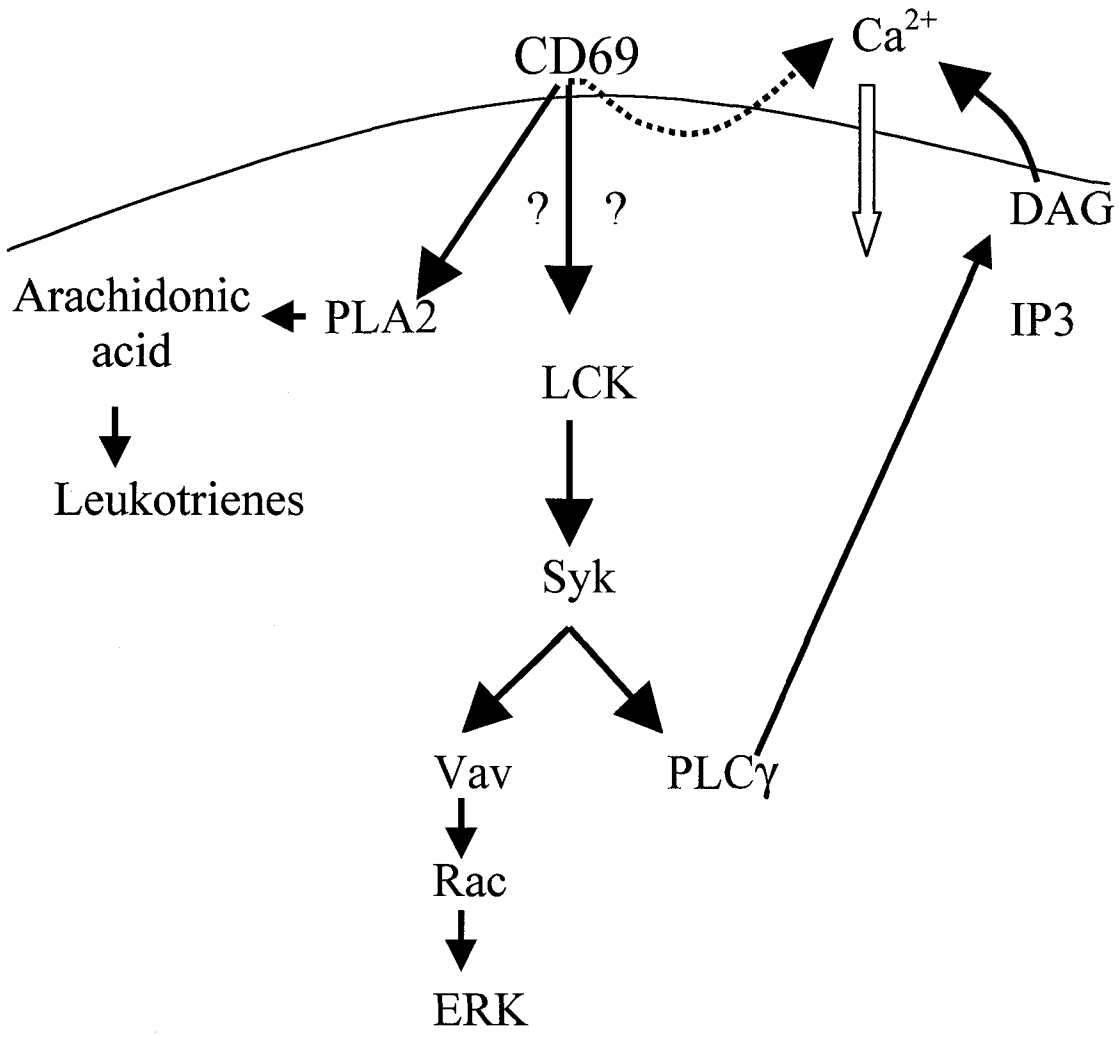


Figure 5: Model of CD69 triggered signaling events. The membrane proximal events associated with CD69 remain unclear. However, it is known that CD69 is capable of signaling via the ERK pathway, phospholipase C and lipid second messengers generated via phospholipase A2 activation.

the calcium induces a slight conformational change allowing access to sites 1 and 2. Site 3 is capable of low-affinity, calcium independent carbohydrate binding (Pavlicek *et al.*, 2003). The residues presumed to be part of ligand binding site one were also involved in binding to an undetermined ligand in the crystal of recombinant soluble CD69 (Natarajan *et al.*, 2000), confirming its probable importance in ligand binding.

1.5.2 AICL

Activation induced C-type lectin (AICL) is a receptor related to (35% amino acid identity) CD69, the AICL gene is also physically near to the gene encoding CD69 within the human NKC (Hamann *et al.*, 1997). Considerably less is know about AICL than CD69 other than the fact that, like CD69, AICL is rapidly expressed on activated leukocytes (Eichler *et al.*, 2001) but is expressed at only very low levels by resting cells. The AICL protein is encoded by a 447 bp open reading frame and is composed of 149 amino acids with a molecular mass of 17.3 kDa before post-translational modifications. Like the related lectin-like receptors, AICL shows a type II transmembrane organization. It features an 18 amino acid transmembrane domain, a seven amino acid cytoplasmic domain, a nine amino acid neck and a C terminal CRD of 115 amino acids. As with CD69, AICL is expressed broadly in different cells of the hematopoietic lineage including lymphocytes, monocytes and granulocytes. Expression of AICL mRNA is detected within one hour with a maximum level of induction at three hours and a decrease in expression at the eighth hour post-stimulation. At the tissue level, AICL is only detectably expressed in the spleen (Hamann *et al.*, 1997). The AICL mRNA is short-lived, with a half-life of less than one hour. AICL mRNA turnover is regulated by a TTTTGTA motif in its 3′ UTR (Hamann *et al.*, 1997).

1.5.3 LLT1

Another related receptor encoded in the human NKC is lectin-like transcript 1 (LLT1). LLT1 is a 191 amino acid receptor featuring a 30 amino acid cytoplasmic domain, a 22 amino acid transmembrane domain and a 132 amino acid CRD. It shows 59% similarity to AICL and 56% similarity to CD69. The LLT1 gene maps near to these genes within the NKC, however, unlike those receptors, LLT1 is preferentially expressed by NK cells. LLT1 is expressed at detectable levels in the spleen and in PBMCs, but not in T or B cell lines (Boles *et al.*, 1999).

A unique feature of LLT1 is a small but significant deviation from the CRD consensus sequence. Most long-form C-type lectin CRDs feature six conserved cysteine residues which form three intrachain disulphide bonds that are important for the proper folding of the CRD; in LLT1 one of these cysteines, involved in the disulphide bond between β strand 3 and the loop following β strand 4 is missing which may lead to a looser conformation of the CRD.

1.5.4 CLR

The C-type lectin-related (CLR) family of genes represents the closest murine homologs of LLT1. Members of the family are 80-90 percent homologous to each other at the nucleic acid level. The seven members of the CLR family are encoded by genes interspersed between the NKRP1 genes and CD69. Expression of CLRB, CLRF and CLRG was detectable by RT-PCR in activated NK cells. The proteins encoded by these three cDNAs are 203-205 amino acids long and 49-67% identical to each other. These proteins are 35-43% identical to AICL, 43-48% identical to LLTI and approximately 40% identical to CD69. Like LLT1, all of the CLR members are missing at least one of

the cysteines conserved in other C-type lectins. Some of the other family members including CLRB and CLRG are also missing the fourth cysteine that would normally pair with the other missing cysteine, cysteine five. Like CD69 and AICL, the CLR genes are organized as five exons rather than six, lacking a discrete neck encoding exon that is present in other NKC encoded lectins (Plougastel *et al.*, 2001).

Each CLR family member shows a different expression pattern. In fact, only CLRB, CLRF and CLRG appear to be expressed at detectable levels in any tissue (Plougastel *et al.*, 2001). CLRB is the most widely expressed family member with transcripts detectable in the spleen, lung, liver, heart, kidney, small intestine and even smooth muscle. CLRF shows a slightly more restricted expression pattern, being found only in the intestine, liver and kidney. The expression of CLRG is even more restricted. Expression of this message was only detectable in NK cells (Plougastel *et al.*, 2001), though it was subsequently shown to be expressed by bone marrow derived macrophages and dendritic cells (Iizuka *et al.*, 2003).

One of the functions attributed to members of the CLR family is to act as ligands for the NKR-P1 family of NK cell receptors (Iizuka *et al.*, 2003). Specifically CLRG, also known as OCILrP2 (Zhou *et al.*, 2002) and DCL1, is a ligand for the activating receptor NKR-P1F. On the other hand, CLRB, also known as OCIL (Zhou *et al.*, 2001), is the ligand for the inhibitory receptor NKR-P1D. It was further demonstrated that CLRB and CLRG are both expressed on bone marrow derived macrophages and dendritic cells, as well as NK cells, and that LPS treatment abolishes expression of both CLRB and CLRG on these cells (Iizuka *et al.*, 2003). Thus, it is likely that CLR/ NKRP1 interactions are physiologically relevant in NK cell-DC interactions. DCs are capable of

activating NK cells in a contact-dependent manner, implying that this occurs via interaction between cell surface molecules such as NKR-P1 and CLR molecules. Activated NK cells are also capable of inducing DC maturation, though this interaction is mediated by a combination of cytokine signaling and cell-cell contact (Gerosa *et al.*, 2002) (Piccioli *et al.*, 2002) (Wilson *et al.*, 1999) (Amakata *et al.*, 2001).

CLRB expression on osteoblasts inhibits osteoclast formation from monocyte-derived precursors (Zhou *et al.*, 2001). Differentiation from monocyte derived osteoclast precursors into functional mutinucleate osteoclasts depends upon cell-to-cell contact between these precursors and osteoblasts or stromal cells, as well as other signals. Thus, it is conceivable that NKR-P1D is expressed by the osteoclast precursors and inhibits differentiation upon recognizing CLRB expressed by the osteoblasts. NKR-P1 has been shown to be expressed on monocytes (Poggi *et al.*, 1997) and thus in all probability in osteoclast precursors as well.

1.6 Natural Killer Cell receptors

1.6.1 NKRP1

NKRP1 genes encode the NK1.1 antigens in mice. Later work showed that NKR-P1 was a member of a family of as many as six genes in mice (Plougastel *et al.*, 2001). In humans, only NKR-P1A has been discovered (Lanier *et al.*, 1994). It is interesting to consider that in rodents, which show expansion of the Ly49 family of lectin-like NK receptors, there is also an apparent expansion of the NKRP1 gene from a single copy to a family as well. However the implications of this beyond a possible general shift to lectin-like receptors in rodents are not clear without more complete knowledge of what role these receptors play in the immune response in both mouse and man.

The first defined members of the family, NKR-P1A (Yokoyama *et al.*, 1991) and NKR-P1C, were shown to be activating receptors, and it was assumed that all members were, though the ligands for these receptors remained unknown. Later it was shown by separate groups that NKR-P1B in mice is functionally an inhibitory receptor (Kung *et al.*, 1999), (Carlyle *et al.*, 1999). However, this fact was originally demonstrated using a cross-linking antibody that triggered inhibitory signals rather than interaction with this receptor's natural ligand.

As mentioned, some murine NKR-P1 family members recognize CLR molecules as their ligands. However, as not all family members were demonstrated to recognize CLR ligands and it remains unclear whether or not LLT1 fulfills an analogous function, thus, it is likely that NKR-P1 members have additional functions that remain to be determined. What little is known about the function of these "NK receptors" is the role they play on other cell types. NKR-P1 is expressed on a subset of T cells and is upregulated upon stimulation with IL-12. It seems that these receptors may be involved in the trans-endothelial migration of these cells to sites of inflammation and injury (Poggi *et al.*, 1999). NKR-P1 molecules are also expressed by monocytes, though the functional significance of this is not clear.

1.7 Chicken group V lectins

In chickens, lectins with homology to AICL, the CLR family and CD69, as well as a gene homologous to NKR-P1, are linked to the MHC genes. The chicken has two MHC encoding regions: the B-locus and the Rfp-Y locus, both of which encode lectin like molecules. Within the B-locus, which contains classical MHC class I and class IIβ genes, are two lectin-encoding genes. One of these, the B-locus lectin (B-Lec) is similar

to the CD69, AICL, LLT1 and the CLR family of genes. The other is B-NK, which bears a cytoplasmic ITIM motif. The expression of B-NK is restricted to chicken NK cells (Kaufman *et al.*, 1999). B-NK shows closest homology to mammalian NKRP1 genes. The Rfp-Y region similarly contains two genes encoding lectin-like receptors. One of these, originally described as gene 17.5 (Bernot *et al.*, 1994), which is likely to be the Y-Lec2 locus (Rogers *et al.*, 2003), is similar to the human AICL gene. Southern blot data, as well as the identification of related transcripts from as yet unidentified loci indicate that this gene is part of a larger gene family (Bernot *et al.*, 1994). Linked to Y-Lec2 within the Rfp-Y region is the gene Y-Lec1, which, like LLT1 and members of the CLR family, lacks one of the conserved cysteines involved in the C-type lectin fold. Y-Lec1 and 2 are localized between the MHC class I-like pseudogenes: Y-FV and Y-FVI (Bernot *et al.*, 1994) (Rogers *et al.*, 2003). No information on the function of these chicken lectins is available.

1.8 Ducks as models for the study of human disease

Ducks are valuable model organisms in the study of hepatitis B virus. They are one of the few animals naturally susceptible to hepatitis B and are the most cost effective model organisms to study this disease. Duck Hepatitis B virus is distinct from the human Hepatitis B virus and they are not cross-infective. However, the pathology of the disease is nearly identical between the two viruses, as is their biology. Various pharmacological treatments have effectively been tested using the duck model, some of which remain in clinical use. However, in the face of rapidly acquired drug resistance, immunotherapies are attractive alternatives for treatment of this disease and are more likely to produce

lasting cures. In this context the weakness of the duck model becomes apparent: the duck immune system is largely uncharacterized.

The biggest obstacle to effective anti-hepatitis B immunotherapy is immune tolerance in the chronically infected individual. Hepatitis B is maintained in the population by chronic carriers, many of which were congenitally infected. This carrier state is associated with immune tolerance to viral antigens. Dendritic cells are incredibly potent professional antigen presenting cells capable of initiating an immune response, inducing peripheral tolerance or even breaking tolerance depending on the context in which they present antigen. The ability to break tolerance to viral antigens was demonstrated using Hepatitis B antigens. Dendritic cells were obtained from transgenic mice expressing Hepatitis B Virus antigens, loaded with HBV antigens in vitro and reintroduced into the host. These DCs were able to induce an immune response towards HBV antigens to which the animal had previously been tolerant (Shimizu *et al.*, 1998). The trials performed in the derived "humanized" mouse model were effective as proof-of-principle, however they are not good indicators of the long-term efficacy of such treatment. It would be useful to duplicate these experiments in a natural model of HBV infection like the duck in order to assess the efficacy of this strategy in inducing a productive immune response in a natural chronic infection. Similar studies showed the same ability to break tolerance against antigens inappropriately expressed by certain tumors (Byrne and Halliday, 2002) or self-antigens abnormally expressed during tumor formation and angiogenesis (Li *et al.*, 2002) that would normally be ignored as "self" antigens, showing that this is not an isolated phenomenon and one well worth exploiting.

Unfortunately, the immune system of ducks remains largely uncharacterized. Dendritic cells have not been identified, let alone receptors to which vaccines could be targeted. Given the abundance and importance of lectin-immunoreceptors they are good candidates for targets of directed vaccine therapy. Thus, identifying duck lectins would not only provide markers to identify cell types or functional states, they would also provide a way to manipulate the activities of these cells. Delivering hepatitis B antigens to DCs in and of itself is not sufficient; these DCs must be primed with the Ag and activated to express a full complement of required co-stimulatory molecules. This could be accomplished by triggering activating lectins such as Dectin-1 and DCAR or perhaps through blocking inhibitory receptors such as DCIR to insure that the vaccine is fully immunogenic.

Similarly, activation antigens such as CD69 could have a dual role in this system if they are identified. These antigens would act as markers of the activated state and as switches to manipulate or enhance the effector activities of immune cells. The findings in the CD69 knockout mouse, for example, demonstrate the importance of CD69 signaling in not only immediate triggering of effector responses but as a mechanism of containing immune activation. Identifying and understanding the role of lectin-like immunoreceptors in ducks will provide us with new tools to manipulate the duck immune response as well as to monitor the effects of these manipulations.

These studies will also provide insight into the evolution of these receptors, which is also very interesting in and of itself. In light of the vastly different repertoires of lectin encoding genes between mice and humans alone, it is clear that the evolution of these

receptors has been rapid and it is difficult to imagine the ancestral state. Examining lectins in another order may provide additional clues as to how these receptors arose.

2 Methods:

2.1 DNA Sequencing

Nucleotide sequencing was performed on both strands using the dyenamic ET terminator cycle sequencing kit (Amersham Pharmacia Biotech) under standard cycle conditions and run on an automated ABI 377 DNA sequencer (PE Applied Biosystems Inc. Foster City, CA). Chromatogram editing, contig assembly, nucleotide alignments and percent identity analysis was performed using the Genetool 2.0 software (Biotools Inc., Edmonton, Alberta) and by visual inspection. Protein characteristics were predicted by analysis with the Peptool 2.0 software (Biotools inc. Edmonton, Alberta). Both Peptool 2.0 and the Net Phos (Blom *et al.*, 1999) and Net N Glyc (Guta *et al.*, 2002) tools of the CBS prediction servers (http://www.cbs.dtu.dk/services/) were used to predict post-translational modifications of these proteins. The structural predictions were made using the Geno 3D prediction servers (Combet *et al.*, 2002) and viewed using the deep view/ Swiss Pdb viewer (Guex and Peitsch, 1997). The gene specific primers used for sequencing are presented in Table 1.

2.2 Phylogenetic analysis

CRD sequences for 84 different lectins from mouse, rat, human, pig, chicken and duck were aligned using the Clustal X software. A bootstrap tree was generated at 10,000 bootstrap trials (Appendix 1). This first trial showed that the duck lectins fell within two distinct subgroups, which were then reanalyzed. A second tree was generated for each group at 1,000 bootstrap trials to clarify the relationships within these subgroups.

2.3 Ducks and tissue collection

All ducks were generously provided by Dr. D. L. Tyrell (University of Alberta) from his colony. White Pekin ducks were euthanized with Euthanol and tissue samples

Table 1:

PRIMER NAME	SEQUENCE	
6D9E1F	CACCCTGACCAAGAAGAAGTAATGA	*
6D93'RACE	GCAAGTGCTACTATTTCTCGGAGG	
6D95' RACE	GGTTTTCCAGTTGCTTTCATCCTCC	*
6D9E4F	ACCACTGGTTCGGGCTGCAC	
29D1UTRF	GTTGCTGTGAGAGGACTGGATGTG	
14D12E1R	CCCATTTGCATTGTTTTTCC	
14D12E3F	CACCCAAATGGATCGGCTTC	
14D12E3F2	TCTCACTCATCACATCCCCAATTCT	
14D12E3R	TTAGAGAAATAGTAGCAC	
14D12E3R2	TCCAGTTGCTTTCATTCTTAGAG	*
14D12E4R	CGTGCAGCCCGAACCAGTGG	
14D12E5F	CAACACGGAGAAATACTGGATC	
14D12E5R	GATCCAGTATTTCTCCGTGTTG	
14B7E1F	CAGAAATCACCTATGCCGAAGT	*
aDCIRE3R	CCAGCTGTCCTCTTTGCCTTGAT	
14B7E4F	GCTATTACTTCTCAGATGATCAG	
14B7E4R	TGCTTCTGTATTGATCACCAC	*
aDCIRE6R	GCTGCAGTCTCACAAATCCGATAAG	
16F11E1F	CGATAATGAACCAGCAAGAGAG	*
16F11E5R	GCCCTCAGACCGATGTATAAATTG	*
βACTINE1F	ACCGCGCAACTCCCCGAAGCCAG	*
βACTINE2R	ATAGCTGTCTTTCTGGCCCATGC	*

Primers marked with an * were used in RT-PCR experiments.

were collected then snap frozen in liquid nitrogen and stored at –80°C for subsequent RNA isolation or in the case of the duodenum processed immediately to obtain intestinal epithelial lymphocytes.

2.4 Cell isolation

To obtain intestinal epithelial lymphocytes (IELs) the duodenum was cut into 1 cm strips and washed three times with cold, sterile PBS. IELs were eluted in PBS supplemented with 0.1mM EDTA and 10mM DTT at 37°C, with shaking for one hour (Gobel, 2000). The supernatant was collected and large sediments were removed by centrifugation for five minutes. The cells were then pelleted by centrifugation for 10 minutes at 100 g and resuspended in 20mL PBS. The cells collected were then separated over Ficoll-Paque, washed and cultured for one hour in RPMI 1640 with 10mM Concanavalin A, 10mM phytohemaglutinin or without stimulation, centrifuged, washed with PBS, pelleted, then snap frozen.

To obtain peripheral blood mononuclear cells (PBMCs) 20 mL whole blood was collected by cardiac puncture immediately following euthanization. The blood was diluted 1:1 in PBS and layered over Ficoll-Paque. The buffy coat layer was collected, washed twice with PBS and snap frozen.

2.5 Southern analysis

High molecular weight DNA extracted from erythrocytes of White Pekin ducks was digested to completion with the restriction enzymes listed. The DNA was then separated on 0.8% agarose gels run overnight at 40V. The DNA was transferred to a Biodyne® B membrane (PALL, East Hills, New York) or Nytran membrane (Schleicher and Schuell, Keene N.H.) and immobilized by UV cross-linking (UV Stratalinker© 2400,

Stratagene, La Jolla, CA). Membranes were pre-hybridized overnight at 42°C in a formamide hybridization solution (50% formamide, 1%SDS, 5% Dextran sulfate, 4X SSPE, 5X Denhardt's reagent). Specific probes were generated by restriction enzyme digestion of clones 14B7 or 14D12 to remove the insert. 14B7 specific probes were also made using primers in exon 1 and exon 3 (apLec 1 E1-3) or in exon 1 and exon 6 (apLec 1 coding), 14D12 specific probes were made using isolated 14D12 plasmid and primers in exon 1 and exon 3 (14D12 E1-3) or in exon 3 and exon 6 (14D12 E3-6). The purified products were then radiolabeled with α^{32}P dCTP by random primer labeling using the Prime It random primer labeling kit (Stratagene, La Jolla, CA) and boiled for 15 minutes with salmon sperm DNA (100μg/mL). The blot was hybridized for 22 hours at 42°C. Low stringency washes were performed using a solution of 1x SSPE, 0.1% SDS with shaking at 52°C. High stringency washes were performed using a 0.1X SSPE, 0.1% SDS solution with shaking at 65°C. The blots were then exposed to Kodak X-OMAT AR Film at –80°C for 8 days.

2.6 Genomic PCR

Genomic DNA was isolated from the erythrocytes of white Pekin ducks. 200 ng genomic DNA from each individual was used as template in separate PCR reactions. To amplify the AICL-like family of lectin-like genes reactions were carried out using 5pmol of the primers 6D9E1F and 14D12E5R (Table 1), which are complementary to all of the members of this group of duck lectin-like genes, in 1.5mM MgCl. PCR was carried out for 35 cycles of a 30 second, 94°C denaturation step, a 30 second, 60°C annealing step and an extension step at 72°C for 105 seconds. All products were cloned into the TOPO 2.1 vector (InVitrogen, Carlsbad, CA) and sequenced as described.

2.7 RNA isolation

Frozen tissue samples were pulverized under liquid nitrogen and RNA was isolated by Trizol extraction (Invitrogen, Carlsbad, CA). Following extraction, the RNA was isopropanol precipitated, washed with 70% ethanol and resuspended in DEPC treated water.

2.8 Northern blotting

Total RNA from each sampled tissue was run on a 0.6% formaldehyde, 1.2% agarose gel and transferred to a Nytran Supercharge nylon transfer membrane (Schleicher and Schuell, Keene N.H.). The membrane was pre-hybridized in a formamide hybridization solution (50% formamide, 1%SDS, 4X SSPE, 5X Denhardt's reagent) and hybridized as above. All probes were prepared as those described above. An initial wash was performed using a 1x SSPE, 0.1% SDS solution at 52°C with two subsequent washes performed at 65°C.

2.9 RT-PCR

First strand cDNA synthesis was performed on 5μg total RNA from various tissues and cell preps using Superscript III (Invitrogen, Carlsbad, CA). 1μL of this first strand cDNA was used as template in gene specific PCR reactions with the primers listed (Table 1). Reaction were carried out for 25 cycles using 5 pmol of each primer, in 1.5mM MgCl using the TaqStart antibody (Clontech, Palo Alto, CA) according to the manufacturers specifications.

First strand cDNA from adherent, non-adherent and LPS treated leukocytes were generously provided by Carrie Chong. PBMCs were obtained as described above. Then $1x10^8$ cells were cultured in a treated non-pyrogenic polystyrene sterile 25 cm^2 tissue

culture flask (Corning Incorporated, Corning, NY) and incubated in 20% v/v bovine serum in RPMI (RPMI-20) at 37 °C. After 24 hours the cells were washed three times with RPMI and the non-adherent cells collected. Both cultures were then incubated for an additional 48 hours in RPMI before RNA isolation.

3 **Results:**

3.1 Identification of cDNAs encoding C-type lectins in the duck spleen

An EST project searching for immunologically relevant genes expressed in the duck spleen uncovered several clones with strong similarity, as shown by a PSI-BLAST search of the Swiss-Prot database, to various genes in the mammalian Natural Killer Complex. These clones fall into two distinct groups (Table 2). One of these groups showed matches to APC associated lectins while the other resembled chicken lectins and mammalian activation induced lectins. All of these clones were sequenced to completion. It was found that three of these clones: 22G6, 29D2 and 33E5 all showed identical coding sequences to 14D12, thus all subsequent analyses were carried out using 14D12 as a representative of this group.

Table 2:

Clone number	Significant matches	E value
6D9	Y-Lec-2/ Gene 17.5	$1e^{-43}$/ $3e^{-37}$
	CLRA	$5e^{-30}$
	Coja Lec2	$2e^{-28}$
	LLT1	1^{-27}
	CD69	$4e^{-26}$
	AICL	$2e^{-25}$
9H3	CLEC-6/CLECSF8	$3e^{-11}$
	MMCL	$1e^{-10}$
	DLEC/BDCA2	$5e^{-10}$
	DCIR	$2e^{-9}$
	FcεRIII	$2e^{-9}$
	Mincle	$7e^{-9}$
14B7	CLECSF8	$5e^{-33}$
	CLECSF11/ DLEC	$4e^{-32}$
	DLEC	$8e^{-32}$
	mMCL	$5e^{-31}$
	DCIR	$5e^{-30}$
	Mincle	$6e^{-29}$
	DCAR	$6e^{-26}$
14D12	Y-Lec-2/ gene 17.5	$1e^{-45}$/ $2e^{-40}$
	CLRA	$7e^{-30}$
	CojaLec2	$6e^{-29}$
	AICL	$4e^{-28}$
	LLT1	$5e^{-28}$
	OCIL	$5e^{-28}$
	CD69	$1e^{-27}$
16F11	CLEC-6/ hsCLECSF8	$7e^{-29}$/ $1e^{-27}$
	mMCL	$4e^{-27}$
	DLEC	$6e^{-25}$
	DCIR	$1e^{-23}$
	DCAR	$2E^{-21}$
22G6	Y-Lec-2/ gene 17.5	$1e^{-45}$/ $2e^{-40}$
	CLRA	$7e^{-30}$
	CojaLec2	$6e^{-29}$
	AICL	$4e^{-28}$
	LLT1	$5e^{-28}$
	OCIL	$5e^{-28}$
	CD69	$1e^{-27}$

29D1	Y-Lec-2/ gene 17.5 CLRA CojaLec2 AICL LLT1 OCIL CD69	$1e^{-45}$/ $2e^{-40}$ $7e^{-30}$ $6e^{-29}$ $4e^{-28}$ $5e^{-28}$ $5e^{-28}$ $1e^{-27}$
33E5	Y-Lec-2/ gene 17.5 CLRA CojaLec2 AICL LLT1 OCIL CD69	$1e^{-45}$/ $2e^{-40}$ $7e^{-30}$ $6e^{-29}$ $4e^{-28}$ $5e^{-28}$ $5e^{-28}$ $1e^{-27}$

Results of BLAST searches performed on the complete sequences of lectin encoding cDNAs identified during the EST project. All E values are derived by PSI-BLAST searches of the Swiss-Prot database

3.2 APC associated lectins in duck

3.2.1 Clone 14B7

The first of these groups represented by clones: 14B7, 14D6 and 16F11. Clone 14B7, was 1117 bases long and encoded a single open reading frame of 711 bases (Figure 6). The start codon was not in a consensus Kozak context (Kozak, 1997); a feature common to many NKC encoded genes. The predicted start codon at position 167 was not the first ATG in the sequence, however the upstream putative start codons were followed by in frame stop codons within 60 nucleotides, indicating that the ATG at 167 is likely to be the true translational start site. This start codon was preceded by a 5′ untranslated region (UTR) of 166 bases and the ORF was followed by a 3′ UTR 219 nucleotides in length featuring a single canonical polyadenylation signal. A second clone, designated 14D6, was found to be identical in sequence to 14B7 but incomplete at the 5′ end.

To determine the type of lectin encoded by 14B7 the amino acid sequence was examined for features typical of C-type lectins encoded in the NKC. The 14B7-encoded polypeptide was 237 amino acids in length and predicted to have a molecular weight of 27.5 kDa before post-translational modifications. 14B7 polypeptide lacked an N-terminal signal sequence but featured a 21 residue putative transmembrane domain, indicating a transmembrane protein of a type II orientation. The N-terminal 42 amino acid cytoplasmic region had a consensus ITIM motif indicating that this protein may function as an inhibitory receptor. The ITIM sequence (ITYAEV) was identical to that found in both human and murine DCIR. This conservation is striking, as there is potential for a great deal of variation within the ITIM consensus of I/VXYXXI/V (Vivier and Daeron, 1997).

```
  1       gtgc ttccagtgt tcagcatgg atgcctttg acagccaaa ggatgaatg
 50  tagaaaagg ttcagcaac tcagtggtg tcttcctct ggatccagg gacctaaag
104  agcaaggcc aagctttta tatcacaaa accaaaaac ggggattgg tgagcacac
158  tgtgattaa atggaagca gaaatcacc tatgccgaa gtgaagttc aagaatgca
                M  E  A   E  I  T   Y  A  E   V  K  F   K  N  A

212  tcaccaact gaagaggtt gaagtacct cagaagaag cagcagcag catgagcaa
      S  P  T   E  E  V   E  V  P   Q  K  K   Q  Q  Q   H  E  Q

266  catacgcag acatgccct ccatggctc ccgtggctg atctcactg ctcctgctc
      H  T  Q   T  C  P   P  W  L   P  W  L   I  S  L   L  L  L

320  ctggtgtgc gttgccctt gttgttgtt ctcctagtc actcacgtc ccccagagc
      L  V  C   V  A  L   V  V  V   L  L  V   T  H  V   P  Q  S

374  tgtgacaag cccgcagtc ctgcagcgg aaccacaca gggtggcac tgcatcttg
      C  D  K   P  A  V   L  Q  R   N  H  T   G  W  H   C  I  L

428  gcagtgcat caaggcaaa gaggacagc tggaagtgc tgtccagag ggctggaga
      A  V  H   Q  G  K   E  D  S   W  K  C   C  P  E   G  W  R
                                          *

482  ccctttcag gaaagctgc tattacttc tcagatgat cagatgccc tggaatgag
      P  F  Q   E  S  C   Y  Y  F   S  D  D   Q  M  P   W  N  E
                     *

536  agcaagaag aactgcagt gggatgggc tcccagctg gtggtgatc aatacagaa
      S  K  K   N  C  S   G  M  G   S  Q  L   V  V  I   N  T  E
                   *

590  gcagagcag gatttcctc tataaggaa ataagaaga cagatgaaa taccaacaa
      A  E  Q   D  F  L   Y  K  E   I  R  R   Q  M  K   Y  Q  Q

644  aatgcaatc aatttattc atcggtctg agggcacag gaggtgggc cagtggcgc
      N  A  I   N  L  F   I  G  L   R  A  Q   E  V  G   Q  W  R

698  tgggcagac cagactccc tataatgaa tcagcagcg ttctggagg tctggggag
      W  A  D   Q  T  P   Y  N  E   S  A  A   F  W  R   S  G  E

752  ccaagtaat aaatctgat gagctgtgt gttgtaatc catcacaaa acagaaaac
      P  S  N   K  S  D   E  L  C   V  V  I   H  H  K   T  E  N
                                *

806  ctccggaac tggaatgat gtcccgtgc agaatacgt tcttatcgg atttgtgag
      L  R  N   W  N  D   V  P  C   R  I  R   S  Y  R   I  C  E
                                *                          *

860  actgcagca gtaactcta tgatggagg aatcctcat cctgagatt agcagcgaa
      T  A  A   V  T  L   *

914  ctgggaaca gcagagggc tgtgttggg aggggtggg agagccttg gagccttta
968  tcttgcctc tgctggtgg gatgatgag actgggagt gatgttgct ctgcacaca
1022 gcatcccct gtgcatgtg tatttctca aagaactat gacgcctca agtagaaac
1076 aataaatgc tagagaact ctgaaaaaa aaaaaaaaa aaaaa
```

Figure 6: Sequence of clone 14B7. Translation of the ORF is shown beneath the nucleotide sequence. The stop codon is indicated by an asterisk. The polyadenylation signal is double underlined. The transmembrane domain is indicated in italics, the ITIM motif is shown in bold and the residues that compose the CRD are underlined. Conserved cysteines are indicated by an asterisk below the residue. Predicted glycosylation sites are boxed.

The amino acid sequence of 14B7 contained a C-type lectin carbohydrate recognition domain at its C-terminus. The long-form CRD, common to many lectin-like immunoreceptors, is characterized by six conserved cysteine residues that form three intrachain disulphide bridges crucial to the three dimensional structure of the CRD. 14B7 featured all six of these cysteines within the predicted CRD. The CRD of 14B7 began with the sequence CCP, a feature common to several dendritic cell and macrophage-associated lectins. Within the CRD of the 14B7 encoded protein was a variant "WIGL" motif of the sequence FIGL. The WIGL motif is common to C-type lectin CRDs; it forms the core of β strand 2 that is important in the proper folding of the domain through mediating interactions between the two hydrophobic cores of the CRD (Zelensky and Gready, 2003). This variant is not unique nor does it represent a significant change in the structure of this essential beta-strand. Classical C-type lectins bind sugar in a calcium dependent manner. Nine residues within the CRD form two calcium-binding sites in classical lectins with site 1 being somewhat dispensable and site 2 being absolutely required for calcium dependent carbohydrate binding. Therefore, it is the six of these nine residues that form site two that are the minimal requirements for a CRD to be capable of calcium-dependent sugar binding. All of the residues that form site 2 were present in the CRD of the mouse and human DCIR as well as the 14B7 encoded protein. Within site 2, a three amino acid motif determines carbohydrate specificity. Both DCIR and the 14B7-encoded protein feature the motif EPS. This motif was more similar to the EPN motif known to confer specificity for mannose than the QPD motif that mediates binding to galactose and related sugars (Drickamer, 1993), however the actual specificity of the EPS motif in vertebrates is unknown.

The CRD is separated from the transmembrane domain by a 36 amino acid neck. The neck region of 14B7 included two cysteine residues that are likely to be involved in interchain disulphide bond formation, as many related receptors are expressed on the cell surface as covalently linked dimers. The presence of more than one cysteine in the neck region of 14B7 is reminiscent of human CLECSF8, which features two cysteines in the neck.

The crystal structures of the CRDs of several lectins have been determined and these known structures including that of mannose binding protein A, CD69 and CD94 were used as templates to model the structure of the 14B7 encoded protein (Figure 7). This analysis revealed a feature unique to the CRD of this group of duck lectins, which is especially pronounced in the 14B7 encoded polypeptide, is an insertion that is predicted to extend the loop between α helix 2 and β strand 2. There is also a similarly extended loop between β strand 3 and β strand 4 in 14B7. Both of these regions have been shown to interact with ligand in other CRDs.

14B7 is predicted to undergo N-linked glycosylation at four sites, three of which are within the CRD. One of these glycosylation sites is conserved between 14B7, human DCIR. This shared glycosylation site may be of functional importance as glycosylation has a role in determining ligand specificity in some lectins (Marshall and Gordon, 2004).

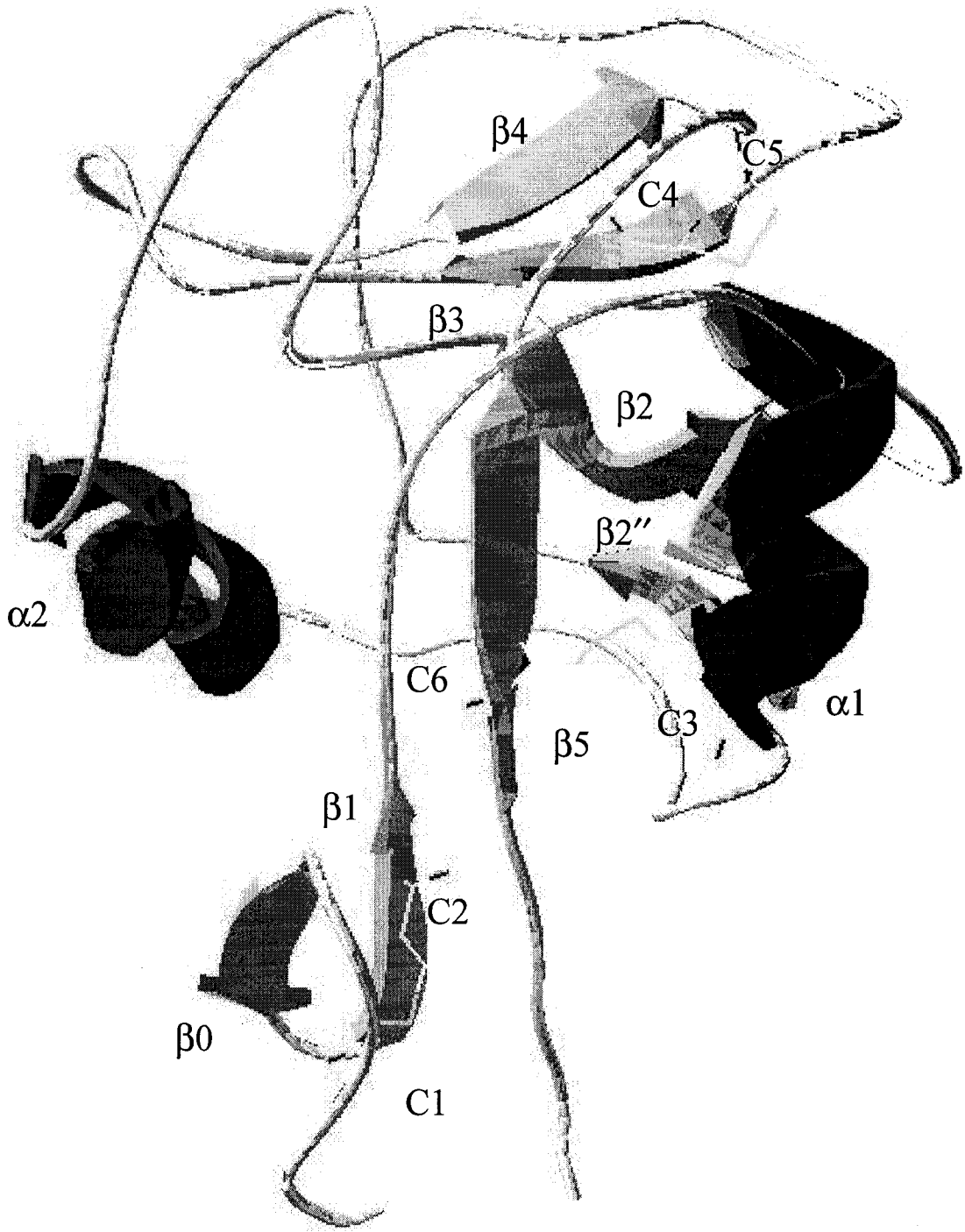


Figure 7: Ribbon diagram of the predicted structure of the 14B7 encoded CRD. Numbering of α helices and β strands is based on that used for MBP-A. The six conserved cysteines forming the three disulphide bonds are shown as stick models and labeled C1-6.

To identify characterized homologs of these genes, the complete sequences were compared to those available in the NCBI database. BLASTN searches with 14B7 showed significant similarity to the gene encoding DCIR. The nucleotide sequence encoding the ORF of clone 14B7 showed 58% identity to murine DCIR and 56% identity to its human homolog, while the mammalian genes show 70% identity to each other. The size of the 14B7 ORF, at 714 base pairs, is comparable to the 714 bp ORF of human DCIR and 717 bp of murine DCIR.

14B7 resembles lectins expressed by antigen presenting cells. Six frame translated BLASTX searches as well as protein-protein BLASTP searches were performed with both sequences to identify homologous proteins and to confirm that the predicted ORF is correct. A BLASTX search with 14B7 cDNA sequence and a BLASTP search with the translation of the 14B7 ORF demonstrated that this protein was most similar to human CLECSF8/MCL (E value BLASTP: $4e^{-32}$; accession number: AAL37713), human DCIR (E value BLASTP: $1e^{-28}$), the N-terminal portion of the rat DCIR-like protein (E value BLASTP: $3e^{-27}$, accession number: XP_232393), and murine DCIR (E value BLASTP: $1e^{-24}$ mouse). All matches with 14B7 were to C-type lectins expressed on APCs therefore we named 14B7 apLec-1 to reflect this similarity to APC lectins.

The amino acid sequence conservation in functionally important regions between apLec-1 and DCIR indicated that these proteins may be functional homologs (Figure 8). Human DCIR, murine DCIR, rat DCIR-like protein, and 14B7 all bear identical ITIM sequences in their cytoplasmic domains. Clone 14B7 and both mouse and human DCIR

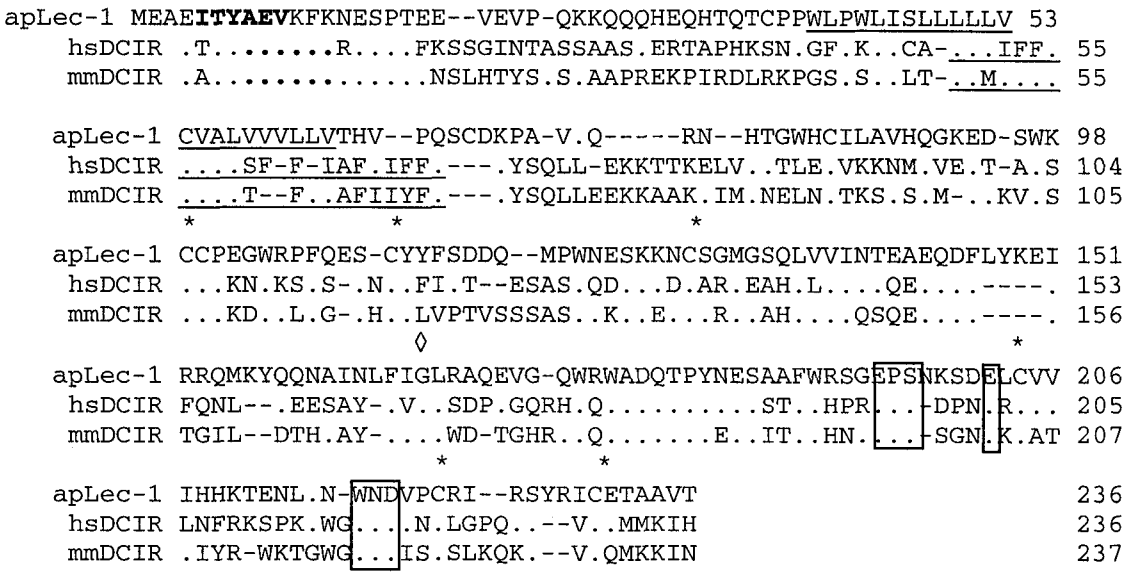


Figure 8: Amino acid alignment of apLec-1 and human and murine DCIR. Dots indicate sequence identity. Dashes indicate gaps introduced to maximize the alignment. Immunoreceptor tyrosine-based inhibitory motif is indicated in bold. Predicted transmembrane regions are underlined. Asterisks represent the positions of the six cysteine residues characteristic of the "long-form CRD". The conserved glycine residue in β strand 2 is indicated by a ◊. The residues required for calcium ligation are in boxes.

also shared 45% sequence identity and approximately 60% homology in their carbohydrate recognition domains. There was little conservation seen in the neck and transmembrane domains, but the cytoplasmic domains of these receptors shared approximately 80% identity and 90% similarity. A BLASTP search with the CRD alone also revealed strong matches to murine MCL (E value: $3e^{-27}$), mMincle (E value: $4e^{-26}$), and BDCA-2 (E value: $4e^{-25}$).

3.2.2 Clone 16F11

16F11 was 1091 base pairs in length, encoding an ORF of 633 base pairs (Figure 9). Like 14B7, the start codon was not in an ideal Kozak context, but the preceding ATG codons were followed by in-frame stop codons. The 5' UTR of 16F11 was 219 base pairs, whereas the 3' UTR was 210 base pairs excluding the poly A tail. 16F11 also featured a single canonical polyadenylation signal.

The 16F11-encoded protein has a distinct, but highly similar CRD (80% identity) that also bears the minimum residues required for calcium binding including the EPS motif seen in 14B7. It also features a region similar to the neck domain of 14B7 and an N-terminal region with some homology to the cytoplasmic region of mammalian APC lectins; however, the 16F11-encoded polypeptide appears to lack a transmembrane region. The16F11 encoded protein also lacked a cleavable N-terminal signal sequence to direct it into the secretory pathway, thus it is likely that 16F11 encoded protein will be located in the cytoplasm. 16F11 had a predicted molecular weight of approximately 24kDa.

```
   1 ggacgggac tcttggtgg agaagtgat gctgtggca gcgataatg aaccagcaa
  55 gagagagtc ggtcctggg actgcaggc tgggggaca gagcccaac taagagttg
 109 ggtagctga gagaaagaa gctccagag agagcccca gcagaaggg agcagctgt
 163 tcccgcctg agcccctgg gtcttcctc gcttctgcc cttgccgtc aaaactgcc
 217 ctcatgacc gttggcctc gttgttctc tttcacagg agctgtggc cagtacaag
       M  T   V  G  L   V  V  L   F  H  R   S  C  G   Q  Y  K

 271 actctgccc cagaatgct ccagggtgg cactgcatc cccaatgga tctgcaagc
     T  L  P   Q  N  A   P  G  W   H  C  I   P  N  G   S  A  S

 325 caaatcact catgtctcc cagagctgt gacaagccc gcagccctg caggggaac
     Q  I  T   H  V  S   Q  S  C   D  K  P   A  A  L   Q  G  N

 379 catacagag ttgcactgc atcttggcg gtgcatcaa ggcaaaggg cgagacctg
     H  T  E   L  H  C   I  L  A   V  H  Q   G  K  G   R  D  L

 433 aagtgctgt tcagagggc tggagaccc tttcaggaa agctgctat tacttctca
     K  C  C   S  E  G   W  R  P   F  Q  E   S  C  Y   Y  F  S
        *                                          *
 487 gatgatcag atgccctgg gatgagagc cagcagaac tgcagtggg atgggctcc
     D  D  Q   M  P  W   D  E  S   Q  Q  N   C  S  G   M  G  S
                                              *
 541 cagctggtg gtgatcaat acaaaagca gagcaggct ttcctctat aaggaaata
     Q  L  V   V  I  N   T  K  A   E  Q  A   F  L  Y   K  E  I

 595 cagatgaaa taccgacaa aatggaatc aatttatac atcggtctg agggcacag
     Q  M  K   Y  R  Q   N  G  I   N  L  Y   I  G  L   R  A  Q

 649 aaggtgggc cagtggcgc tgggcagac cagactccc tataatgaa agagcagcg
     K  V  G   Q  W  R   W  A  D   Q  T  P   Y  N  E   R  A  A

 703 ttctggagg cgtggggag ccaagtgat caaccaagt gatgagctg tgtgttgta
     F  W  R   R  G  E   P  S  D   Q  P  S   D  E  L   C  V  V
                                                    *
 757 atccattac cagaaagat attttccgg aactggaat aatgtccca tgcacaatc
     I  H  Y   Q  K  D   I  F  R   N  W  N   N  V  P   C  T  I
                                                    *
 811 cactcttat tggatttgt gagactgca gcagaaaca atatgatgg aggaatcct
     H  S  Y   W  I  C   E  T  A   A  E  T   I  *
                    *
 865 catcctgag atgagcagc gaactggga acagcagag ggctgtgtt gggaggggt
 919 aggagagcc ttggagcat ttgtcttgc ctctgctcg tgggatgat gagactggg
 973 agtgatgtt gctctgcac acagcatcc cttgtgcat gtgtatttc tcaaagtac
1027 ccctgatgt ggaaacaat aaatgctag agaactctg aaaaaaaaa aaaaaaaaa
1081 aaaaaaaaa aa
```

Figure 9: Sequence of clone 16F11. Translation of the ORF is shown beneath the nucleotide sequence. The stop codon is indicated by an asterisk. The polyadenylation signal is double underlined. The residues that compose the CRD are underlined. Conserved cysteines are indicated by an asterisk below the residue.

There is precedence for transmembrane deletion variants of mRNAs encoding lectin-like receptors, including the DCIR gene (Huang *et al.*, 2001) (Richard *et al.*, 2002). To determine if the 16F11 cDNA represented such a transmembrane deletion variant, or if this clone represented the only form of this message, RT-PCR experiments were conducted using a forward primer based on a unique sequence in the 5′ UTR and a reverse primer in the CRD encoding region to produce products that would span the site of a potential transmembrane encoding exon. This PCR generated three products, one of these was 548 bp, one was 451 bp and the third was 466 bp (Figure 10a). These products were run on an agarose gel, purified, cloned into the Topo 2.1 vector and sequenced.

Interestingly this analysis revealed that clone 16F11 was incorrectly spliced to include intron sequences that altered the reading frame. The largest product appears to be the correctly spliced form, and was called apLec-2. It featured a larger open reading frame that included a transmembrane domain-encoding region and used the same CRD encoding sequence as that in the original library clone. This transmembrane region included a positively charged residue, which may be involved in associations with adaptor molecules in a manner similar to activating immunoreceptors. The 451 bp transcript, apLec-2b, lacked a portion of exon 2 that eliminated some of the hydrophobic residues involved in the transmembrane domain, leaving an insufficient length of hydrophobic residues to form a membrane spanning domain. This variant is especially similar to the transmembrane deletion variants seen in DCIR (Huang *et al.*, 2001) (Richard *et al.*, 2002). The 466 bp variant, apLec-2c, is spliced such that it encodes a unique neck region but like apLec-2b lacks a transmembrane domain. The

A

B

```
  >16-F11 GGACGGGACTCTTGGTGGAGAAGTGATGCTGTGGCAGCGATAATGAACCAGC 52
 >apLec-2                                          CGATAATGAACCAGC 15
>apLec-2b                                          CGATAATGAACCAGC 15
>apLec-2c                                          CGATAATGAACCAGC 15

  >16-F11 AAGAGAGAGTCGGTCCTGGGACTGCAGGCTGGGGGACAGAGCCCAACTAAGA 104
 >apLec-2 AAGAGAGAGTCGGTCCTGGGACTGCA-------------------------- 41
>apLec-2b AAGAGAGAGTCGGTCCTGGGACTGCA-------------------------- 41
>apLec-2c AAGAGAGAGTCGGTCCTGGGACTGCA-------------------------- 41

  >16-F11 GTTGGGTAGCTGAGAGAAAGAAGCTCCAGAGAGAGCCCCAGCAGAAGGGAGC 156
 >apLec-2 --------------------------------GCCCCAGCAGAAGGGAGC 59
>apLec-2b ---------------------------------------------------- 41
>apLec-2c ---------------------------------------------------- 41

  >16-F11 AGCTGTTCCCGCCTGAGCCCCTGGGTCTTCCTCGCTTCTGCCCTTGCCGTCA 208
 >apLec-2 AGCTGTTCCCGCCTGAGCCCCTGGGTCTTCCTCGCTTCTGCCCTTGCCGTCA 111
>apLec-2b ---------------------------------------------------- 41
>apLec-2c ---------------------------------------------------- 41

  >16-F11 AAACTGCCCTCATGACCGTTGGCCTCGTTGTTCTCTTTCACAGGAGCTGTGG 260
 >apLec-2 AAACTGCCCTCATGACCGTTGGCCTCGTTGTTCTCTTTCACAGGAGCTGTGG 163
>apLec-2b ------------------------GTTGTTCTCTTTCACAGGAGCTGTGG 67
>apLec-2c ------------------------------------------GGAGCTGTGG 51

  >16-F11 CCAGTACAAGACTCTGCCCCAGAATGCTCCAGGGTGGCACTGCATCCCCAAT 312
 >apLec-2 CCAGTACAAGACTCTGCCCCAGAATGCTCCAGGGTGGCACTGCATCCCCAAT 215
>apLec-2b CCAGTACAAGACTCTGCCCCAGAATGCTCCAGGGTGGCACTGCATCCCCAAT 119
>apLec-2c CCAGTACAAGACTCTGCCCCAGAATGCTCCAGGGTGGCACTGCATCCCCAAT 103

  >16-F11 GGATCTGCAAGCCAAATCACTCATGTCTCCCAGAGCTGTGACAAGCCCGCAG 364
 >apLec-2 GGATCTGCAAGCCAAATCACTCATGTCTCCCAGAGCTGTGACAAGCCCGCAG 267
>apLec-2b GGATCTGCAAGCCAAATCACTCATGTCTCCCAGAGCTGTGACAAGCCCGCAG 171
>apLec-2c GGATCTGCAAGCCAAATCACTCATGTCTCCCAGAGCTGTGACAAGCCCGCAG 155

  >16-F11 CCCTGCAGGGGAACCATACAGAGTTGCACTGCATCTTGGCGGTGCATCAAGG 416
 >apLec-2 CCCTGCAGGGGAACCATACAGAGTTGCACTGCATCTTGGCGGTGCATCAAGG 319
>apLec-2b CCCTGCAGGGGAACCATACAGAGTTGCACTGCATCTTGGCGGTGCATCAAGG 223
>apLec-2c CCCTGCAGGGGAACCATACAGAGTTGCACTGCATCTTGGCGGTGCATCAAGG 207

  >16-F11 CAA------------------------------AGGGCGAGACCTGAAGTG 437
 >apLec-2 CAA------------------------------AGGGCGAGACCTGAAGTG 340
>apLec-2b CAA------------------------------AGGGCGAGACCTGAAGTG 244
>apLec-2c CAAAGCATTCATGCTCCGTGGTCTCTCCCTGCACAGGGCGAGACCTGAAGTG 259
```

```
  >16-F11 CTGTTCAGAGGGCTGGAGACCCTTTCAGGAAAGCTGCTATTACTTCTCAGAT 489
 >apLec-2 CTGTTCAGAGGGCTGGAGACCCTTTCAGGAAAGCTGCTATTACTTCTCAGAT 392
>apLec-2b CTGTTCAGAGGGCTGGAGACCCTTTCAGGAAAGCTGCTATTACTTCTCAGAT 296
>apLec-2c CTGTTCAGAGGGCTGGAGACCCTTTCAGGAAAGCTGCTATTACTTCTCAGAT 311

  >16-F11 GATCAGATGCCCTGGGATGAGAGCCAGCAGAACTGCAGTGGGATGGGCTCCC 541
 >apLec-2 GATCAGATGCCCTGGGATGAGAGCCAGCAGAACTGCAGTGGGATGGGCTCCC 444
>apLec-2b GATCAGATGCCCTGGGATGAGAGCCAGCAGAACTGCAGTGGGATGGGCTCCC 348
>apLec-2c GATCAGATGCCCTGGGATGAGAGCCAGCAGAACTGCAGTGGGATGGGCTCCC 363

  >16-F11 AGCTGGTGGTGATCAATACAAAAGCAGAGCAGGCTTTCCTCTATAAGGAAAT 593
 >apLec-2 AGCTGGTGGTGATCAATACAAAAGCAGAGCAGGCTTTCCTCTATAAGGAAAT 496
>apLec-2b AGCTGGTGGTGATCAATACAAAAGCAGAGCAGGCTTTCCTCTATAAGGAAAT 400
>apLec-2c AGCTGGTGGTGATCAATACAAAAGCAGAGCAGGCTTTCCTCTATAAGGAAAT 415

  >16-F11 ACAGATGAAATACCGACAAAATGGAATCAATTTATACATCGGTCTGAGGGCA 645
 >apLec-2 ACAGATGAAATACCGACAAAATGGAATCAATTTATACATCGGTCTGAGGGC  547
>apLec-2b ACAGATGAAATACCGACAAAATGGAATCAATTTATACATCGGTCTGAGGGC  451
>apLec-2c ACAGATGAAATACCGACAAAATGGAATCAATTTATACATCGGTCTGAGGGC  466

  >16-F11 CAGAAGGTGGGCCAGTGGCGCTGGGCAGACCAGACTCCCTATAATGAAAGAG 697

  >16-F11 CAGCGTTCTGGAGGCGTGGGGAGCCAAGTGATCAACCAAGTGATGAGCTGTG 749

  >16-F11 TGTTGTAATCCATTACCAGAAAGATATTTTCCGGAACTGGAATAATGTCCCA 801

  >16-F11 TGCACAATCCACTCTTATTGGATTTGTGAGACTGCAGCAGAAACAATATGAT 853

  >16-F11 GGAGGAATCCTCATCCTGAGATGAGCAGCGAACTGGGAACAGCAGAGGGCTG 905

  >16-F11 TGTTGGGAGGGGTAGGAGAGCCTTGGAGCATTTGTCTTGCCTCTGCTCGTGG 957

  >16-F11 GATGATGAGACTGGGAGTGATGTTGCTCTGCACACAGCATCCCTTGTGCATG 1009

  >16-F11 TGTATTTCTCAAAGTACCCCTGATGTGGAAACAATAAATGCTAGAGAACTCT 1061

  >16-F11 GAAAAAAAAAAAAAAAAAAAAAAAAAA                         1090
```

C

```
  16-F11                                                MTVGLVVLFHRSCG 14
  apLec-2 MNQQERVGPGTAAPAEGSSCSRLSPWVFLASALAVKTALMTVGLVVLFHRSCG 53
 apLec-2b MNQQERVGPGTA------------------------------VVLFHRSCG 21
 apLec-2c MNQQERVGPGTA--------------------------GAVASTRLCPRMLQ 26

   16-F11 QYKTLPQNAPGWHCIPNGSASQITHVSQSCDKPAALQGNHTELHCILAVHQGK 67
  apLec-2 QYKTLPQNAPGWHCIPNGSASQITHVSQSCDKPAALQGNHTELHCILAVHQGK 106
 apLec-2b QYKTLPQNAPGWHCIPNGSASQITHVSQSCDKPAALQGNHTELHCILAVHQGK 74
 apLec-2c GGTASPMDLQAKSLMSPRAVTSPQPCRGTIQSCTASWRCIKAKHSCSVVSPCT 79

   16-F11 GRDLKCCSEGWRPFQESCYYFSDDQMPWDESQQNCSGMGSQLVVINTKAEQAF 120
  apLec-2 GRDLKCCSEGWRPFQESCYYFSDDQMPWDESQQNCSGMGSQLVVINTKAEQAF 159
 apLec-2b GRDLKCCSEGWRPFQESCYYFSDDQMPWDESQQNCSGMGSQLVVINTKAEQAF 127
 apLec-2c GRDLKCCSEGWRPFQESCYYFSDDQMPWDESQQNCSGMGSQLVVINTKAEQAF 132
```

```
  16-F11 LYKEIQMKYRQNGINLYIGLRAQKVGQWRWADQTPYNERAAFWRRGEPSDQPS 173
 apLec-2 LYKEIQMKYRQNGINLYIGLRAQKVGQWRWADQTPYNERAAFWRRGEPSDQPS 212
apLec-2b LYKEIQMKYRQNGINLYIGLRAQKVGQWRWADQTPYNERAAFWRRGEPSDQPS 180
apLec-2c LYKEIQMKYRQNGINLYIGLRAQKVGQWRWADQTPYNERAAFWRRGEPSDQPS 185

  16-F11 DELCVVIHYQKDIFRNWNNVPCTIHSYWICETAAET                   209
 apLec-2 DELCVVIHYQKDIFRNWNNVPCTIHSYWICETAAET                   248
apLec-2b DELCVVIHYQKDIFRNWNNVPCTIHSYWICETAAET                   216
apLec-2c DELCVVIHYQKDIFRNWNNVPCTIHSYWICETAAET                   221
```

Figure 10: The gene encoding 16F11 undergoes alternative splicing. A: RT-PCR experiments using primers specific for clone 16F11 yielded three different products. Lane 1 corresponds to reactions carried out with duck 26 spleen cDNA as template. Lane 2 is the negative control. B: Nucleotide alignment of the library clone 16F11 and the sequences of the three cloned RT-PCR products generated with primers specific for 16F11. C: Alignment of the amino acid sequence of each of the forms of the apLec-2.

frequency of differentially spliced messages is in part due to the poor splice and acceptor sites at this intron exon boundary. Neither the correctly spliced form, the alternately spliced form nor the incorrectly spliced library clone make use of canonical donor splice site (AG/GTRAGT) or acceptor splice site (YYTTYYYYYYNCAG/G) this intron. Interestingly, none of the RT-PCR products were identical to 16F11, suggesting that this form of the transcript is either extremely rare or was an artifact of the library construction.

A sequence was assembled from the amplified product and the 3′ portion of clone 16F11 to reflect what is likely to be the correctly spliced form of apLec-2 (Figure 11). ApLec-2 features four distinct domains: a short, N-terminal cytoplasmic domain, a transmembrane region and a neck separating the C-terminal CRD from the transmembrane region. The cytoplasmic domain of the 16F11 encoded protein, like many NKC encoded activating receptors, lacked intrinsic signaling motifs. Thus, any signaling capacity of this receptor is likely to be achieved by association with adaptor molecules via the positively charged residue in its transmembrane domain. The neck region of 16F11 is large, at 61 amino acids, and, like 14B7, featured several cysteines, which may mediate disulphide bond formation between members of a dimer. The CRD of 16F11 was 80.4% identical to that of 14B7 at the amino acid level. Among the shared features was the conservation of the residues that compose calcium binding site two, including the EPS motif in the carbohydrate specificity-determining region. As is the case in apLec-1, the apLec-2 CRD had amino acid insertions that extend the loop between α helix 2 and β strand 2, although this insertion is two amino acids shorter in 16F11 than it is in 14B7.

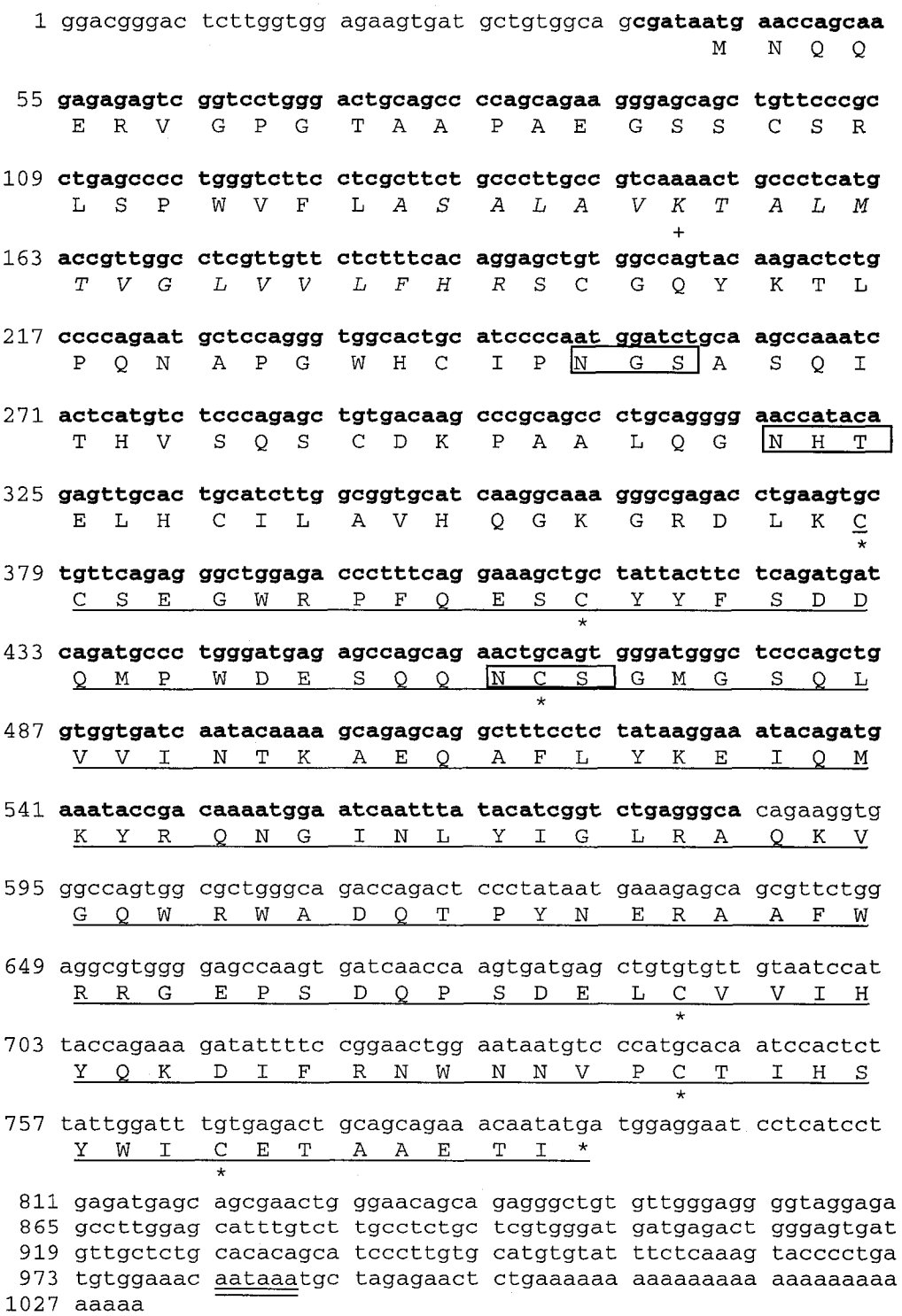


Figure 11: Contig sequence assembled from clone 16F11 and the transmembrane domain encoding RT-PCR product. Nucleotide sequence obtained from the RT-PCR product is shown in bold. Translation of the ORF is shown beneath the nucleotide sequence. The stop codon is indicated by an asterisk. The polyadenylation signal is double underlined. The transmembrane domain is indicated in italics, the positively charged residue in this region is indicated by a +, and the residues that compose the CRD are underlined. Conserved cysteines are indicated by an asterisk below the residue. Predicted glycosylation sites are boxed.

The predicted structure of the apLec-2 CRD (Figure 12) reveals other subtle differences between the CRD of this lectin compared to apLec-1. One distinguishing feature in addition to the slightly shorter loop between α2 and β2 is the extension of the β3 strand. While this extension may be an artifact of the model and not represented in the actual structure of this CRD it contrasts strongly with the short β3 strand and extended loop between B3 strand and β4 strand seen in the model of apLec-1. As these are regions that are likely to contact ligand, the differences between them may indicate that these two similar proteins bind distinct ligands.

Like apLec-1, the apLec-2 encoded protein is predicted to undergo extensive glycosylation. There are three sites predicted to undergo N-linked glycosylation, with only one of these sites within the CRD.

BLAST analysis performed with the sequence of 16F11 failed to identify any single potential homolog. ApLec-2 showed matches with very similar E values to human CLECSF8/MCL (E value: $1e^{-26}$), Macrophage-inducible C-type Lectin (Mincle) (E value: $3e^{-26}$), DCIR (E value: $2e^{-24}$ (rat homolog) $6e^{-21}$ (human)), mouse MCL (E value: $1e^{-22}$) and matches with higher, but still significant, E values to other APC associated lectins.

3.2.3 apLec-2 and apLec-1 are an activating/inhibitory receptor pair

Clustal X analysis was performed to further clarify the relationship between these duck proteins and various mammalian lectins. The phylogenetic tree derived from this alignment indicates apLec-1 and apLec-2 are more closely related to each other than to any given mammalian receptor, however, the pair does fall within a family of lectins expressed on APCs including DCIR (Figure13). Taken together these results indicate

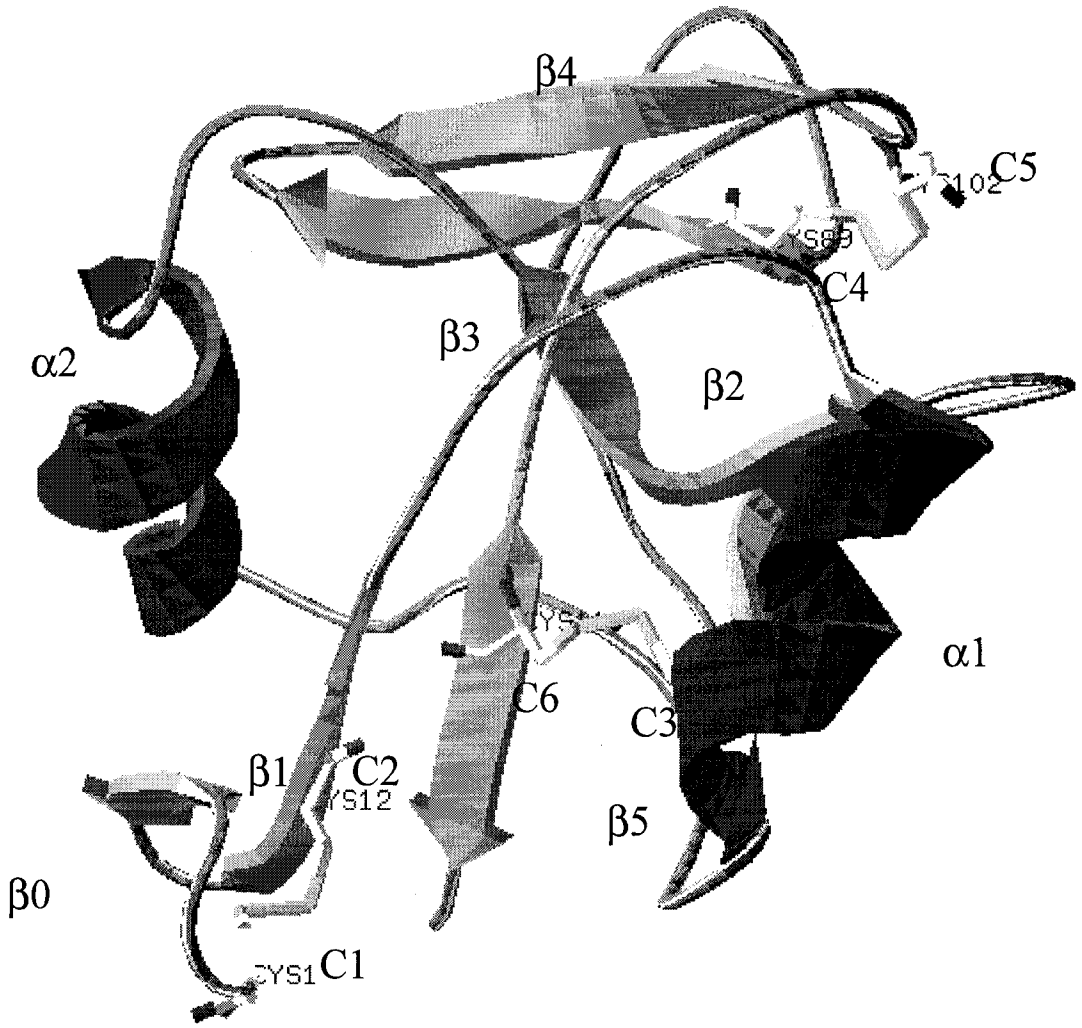


Figure 12: Ribbon diagram of the predicted structure of the 16F11 encoded CRD. Numbering of secondary structural features is based on that used for MBP-A. The six conserved cysteines (C1-6) and the three disulphide bonds they form are shown as stick models.

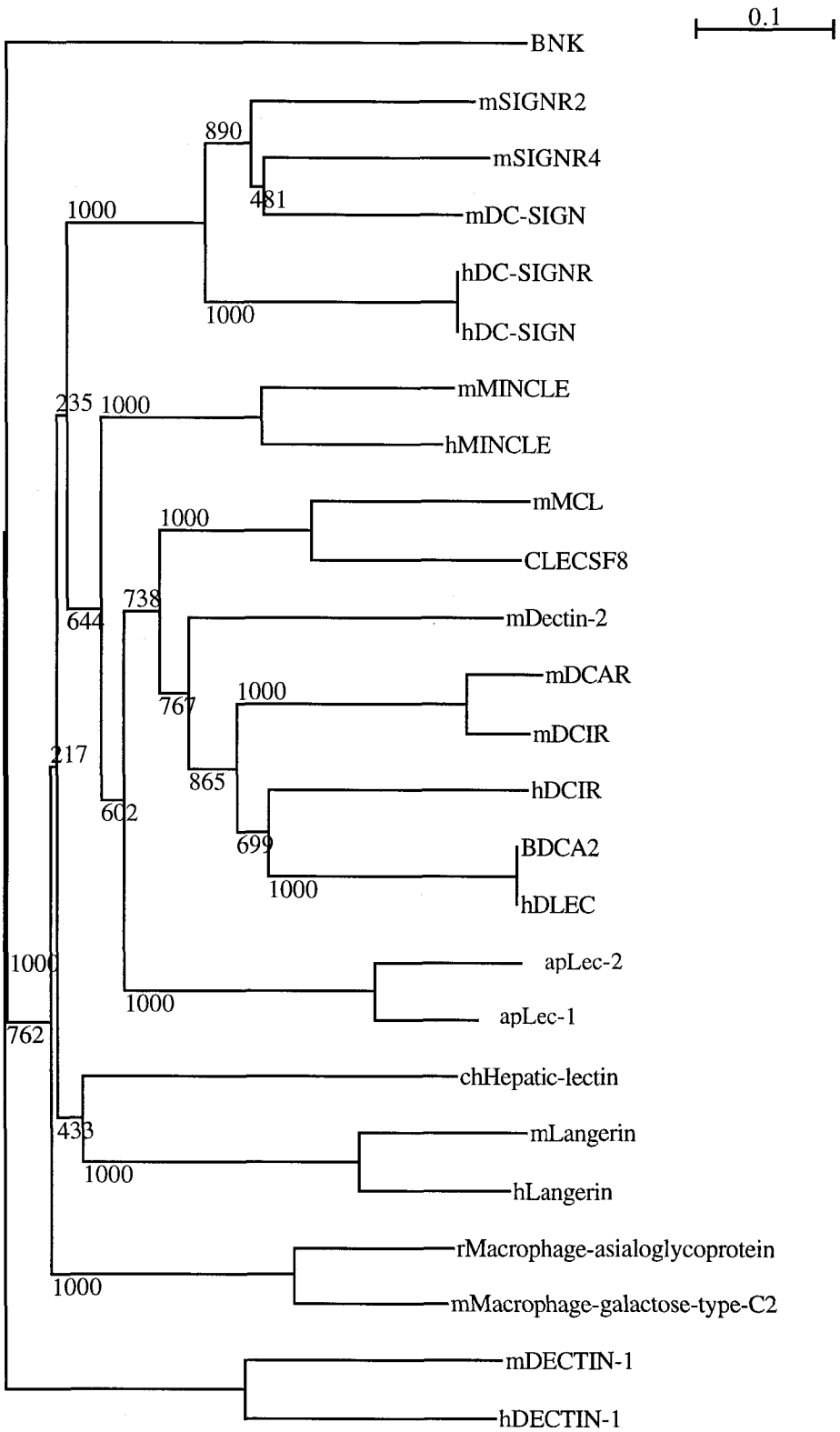


Figure 13: Relationship between the apLec proteins and APC associated lectins. Dendrogram generated based on CRD sequences of the novel duck proteins apLec-1, apLec-2, and various APC associated lectins generated using the Clustal X program and NJ plot. The chicken NK cell receptor BNK was included to root the tree.

that apLec-1 and apLec-2 represent duck homologs of mammalian APC associated lectin-like receptors.

The degree of similarity between the CRDs of this protein and the one encoded by apLec1 suggests they may in fact form an activating and inhibitory receptor pair and that apLec-2 may represent an activating counterpart to apLec-1 in a manner analogous to DCIR and DCAR in mice (Kanazawa *et al.*, 2003). ApLec-1 and apLec-2, were found to be approximately 81% identical at the cDNA level (Appendix 2) and encoded proteins that shared 67% identity in amino acid sequence. An alignment of the amin acid sequences revealed that this similarity extends to the EPS motif in the carbohydrate specificity determining pocket of Ca^{2+} binding site 2 and the extended loops described in the CRD of 14B7, although a deletion of two amino acids in the 16F11 CRD reduces the size of this loop slightly and alters its predicted fold (Figure 14).

ApLec-1 and apLec-2 showed very little conservation in their cytoplasmic domains, which is unsurprising as the cytoplasmic domains of inhibitory receptors are quite distinct from those of activating receptors. Similarly, since there is no selective pressure to conserve the sequence of the transmembrane domains there was a great deal of divergence between this part of these receptors as well. Of central importance was the presence of a positively charged residue in the TM of apLec-2, necessary for it to function as an activating receptor, and the absence of any charged residues in the TM of apLec1. The neck region of apLec2 is quite large, a full 27 amino acids longer than that of apLec1. The neck region of apLec-1 shared 74.3% identity with the C-terminal portion of the neck of apLec-2. With one exception, the residues that are not conserved between the stalk regions of the two receptors represent substantial changes such as the

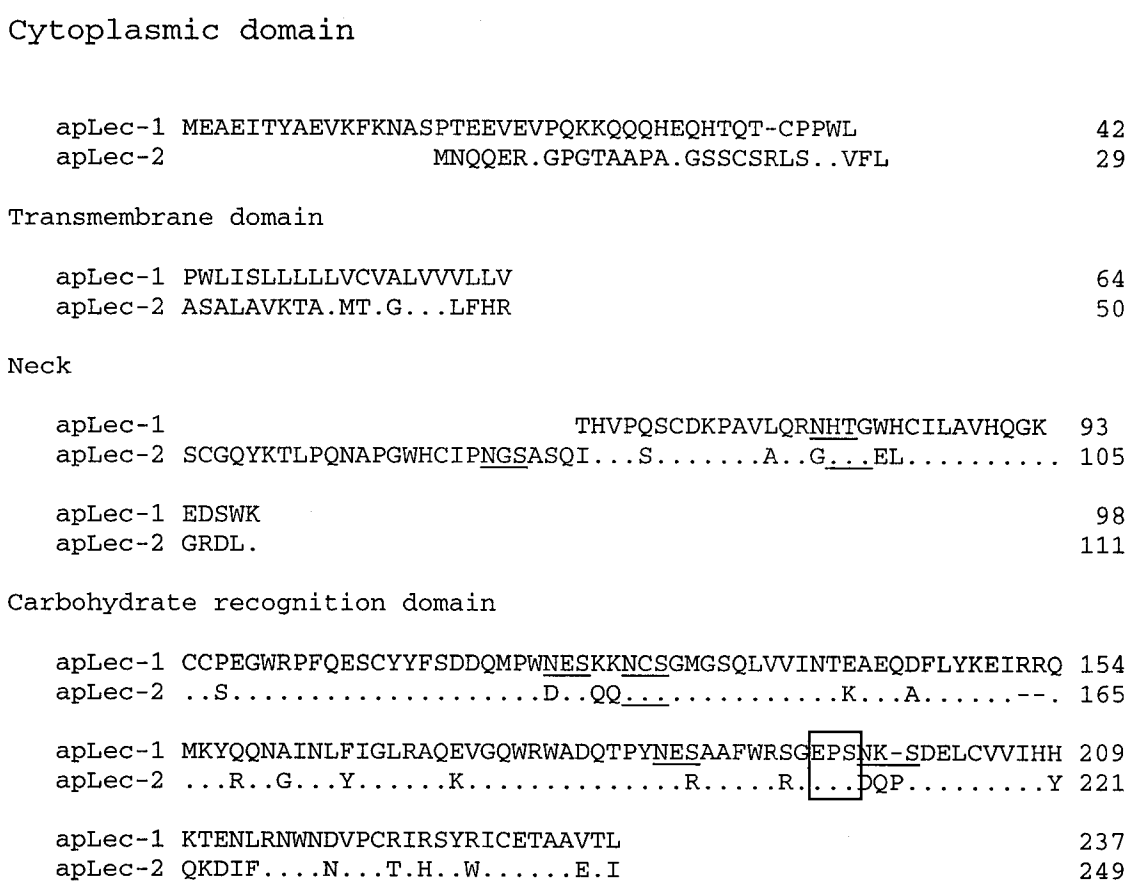


Figure 14: Amino acid alignment of apLec-1 and apLec-2. In both alignments, dots indicate identical residues and dashes indicate gaps introduced to maximize the alignments. The EPS motif shared by these receptors is boxed, predicted glycosylation sites are underlined.

bulky, positively charged arginine at position 78 of apLec-1 being substituted by a small, non-polar glycine at the equivalent position of apLec-2. The changes in the CRD may be significant to either the folding or the ligand specificity of these receptors. Amongst these, there are four positively charged lysines at positions 125, 126, 199 and 210 in apLec-1 being replaced by polar glutamines at equivalent positions in apLec-2. The changes at positions are both in loops that may contact ligand and thus these differences may alter the ligand binding properties of these CRDs. There were also two cases where positively charged residues in apLec-1 were replaced with hydrophobic residues in apLec-2. One of these, a histidine at position 209 in apLec-1 was replaced with a tyrosine at the equivalent position in apLec-2 may be especially significant. This substitution coupled with the lysine to glutamine change at position 210 result in three positively charged residues in a row in apLec-1 being replaced by a positively charged residue a hydrophobic residue and a polar residue in apLec-2. These changes in the loop between the β 3 strand and β 4 strand could influence ligand binding. There are also two cases where a negatively charged glutamic acid residue in apLec-1 is replaced by a positively charged lysine in apLec-2, this occurs at positions 141 and 172. Such differences suggest that these CRDs will have different binding characteristics, due to the differences in charges and conformation of the CRD induced by these substitutions. These sequence differences also result in very different glycosylation patterns between apLec-1 and apLec2. The CRD of apLec-1 is N-glycosylated at as many as four sites within the CRD wheras apLEc-2 ony has one N-glycosylation site. Thus at minimum if these receptors bind the same ligand they can be expected to bind with different affinities, but it seems even more likely that they bind distinct but possibly related ligands.

3.2.4 apLec-1 and apLec-2 are encoded by one or two copy genes.

To determine the copy number of the genes encoding apLec-1 and apLec-2, duck genomic DNA was digested using a variety of enzymes and transferred to a membrane (Figure 15A). This blot was then hybridized with a probe starting from the coding portion of exon one and ending in the neck domain encoding exon three of clone 14B7 (apLec-1 E1-3) (Figure 15A). This probe excludes the CRD encoding portion of the cDNA, to avoid non-specific hybridization to other lectin-encoding genes. The blot was probed under both high and low stringency conditions. This revealed either single bands or a simple restriction pattern depending on the restriction enzyme used, indicating that apLec-1 exists as a single copy gene in the duck genome.

Similar experiments were performed using a probe derived from apLec-2 cDNA, also designed to exclude the majority of the CRD encoding region (Figure 15B). The results suggests one or two genes based on the simple restriction pattern. The number of identical bands that appear on the two blots imply that either the apLec-1 and apLec-2 genes are linked and fall on the same restriction fragment or that these genes arose by duplication and that several restriction sites were also duplicated resulting in very similar pattern for these two genes. These possibilities are not necessarily mutually exclusive as the genes may fall on different restriction fragments of identical size and still be linked. The degree of sequence similarity between the two genes support the notion that they arose by duplication, with the differences between the first three presumed exons due to divergence by exon shuffling.

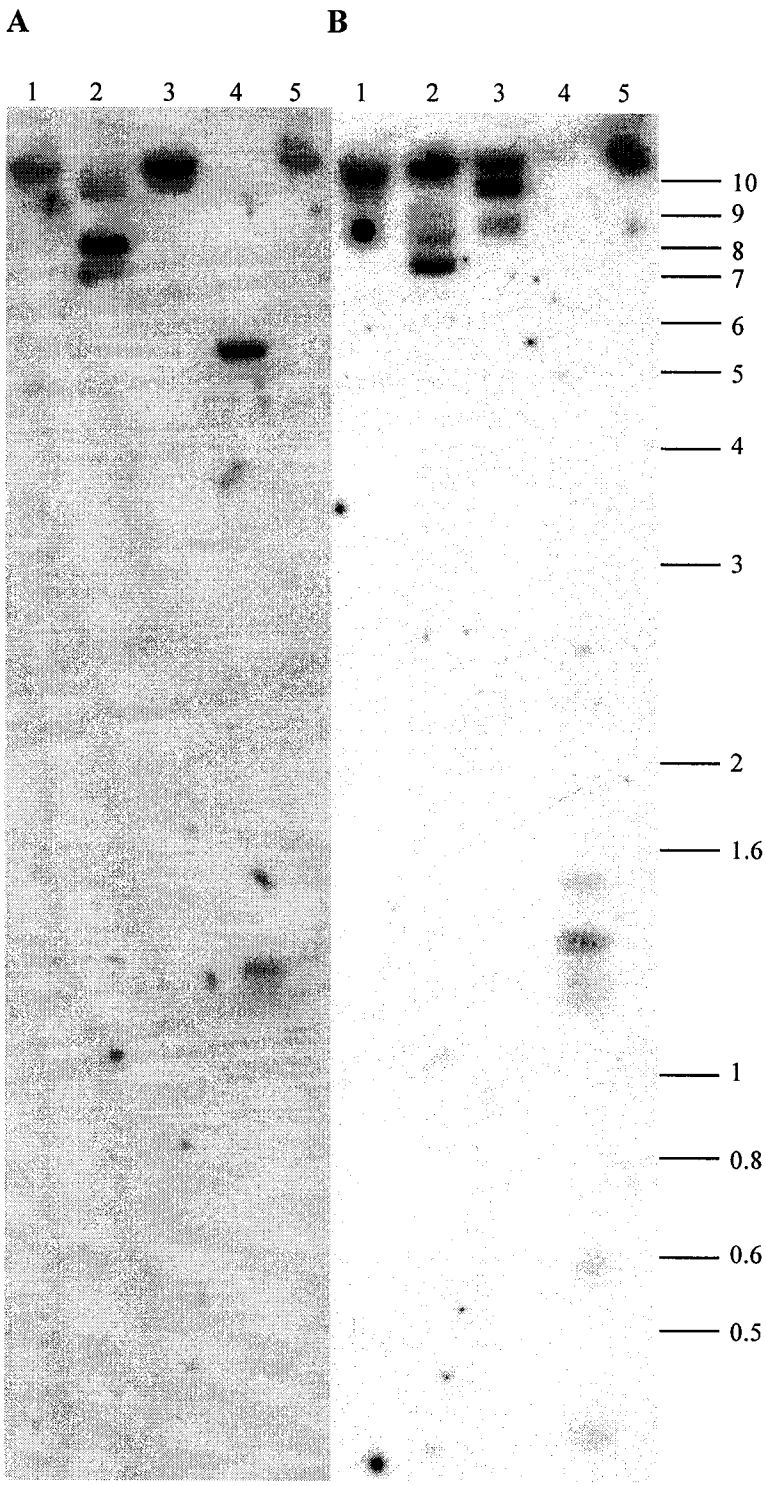


Figure 15: The apLec proteins are encoded by one or two genes. Genomic DNA was digested with several restriction endonucleases and transferred to a Biodyne B® membrane. Lanes correspond to: 1: Xho1, 2:XbaI, 3: SpeI, 4: PvuII, and 5: NotI. Southern blots hybridized with probe derived from A) the portion of 14B7 encoding the cytoplasmic, transmembrane and neck region and B) from immediately upstream of the start codon to the neck encoding portion of the 16F11 cDNA reveals a simple pattern of banding indicating the existence of one or two hybridizing genes. The 1 kb plus ladder (InVitrogen) was used as a size standard.

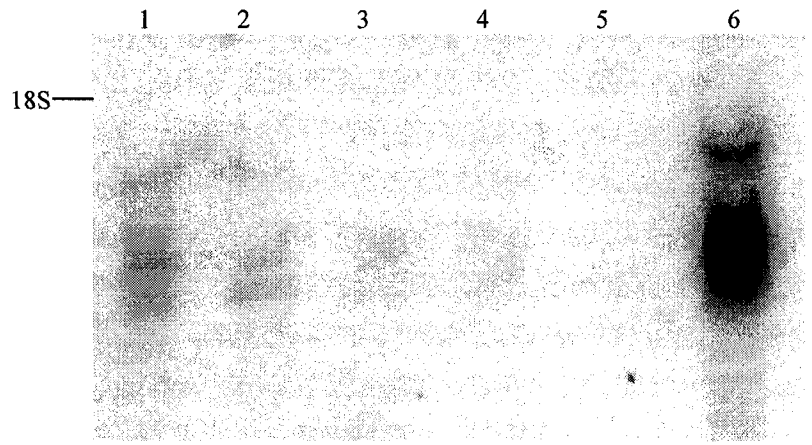
3.2.5 The apLec genes were expressed mainly in immune and mucosal tissues

To determine where apLec-1 and apLec-2 were expressed a northern blot was prepared using total RNA extracted from several tissues of one duck that was congenitally infected with duck hepatitis B virus. This blot was then hybridized with the probe apLec-1 E1-3 and the blot was exposed to film for three days at -80°C. Bands hybridizing to the apLec-1 derived probe were detected in the spleen, intestine and lung (Figure 16A). The same blot was stripped then hybridized with a probe derived from the equivalent region of apLec-2. This blot was exposed for seven days before bands were visible, suggesting that apLec-2 messages was less abundant than apLec-1 mRNA. Bands hybridizing to apLec-2 mRNA were detectable in the spleen, a faint hybridizing band in the lung and an extremely faint band in the intestine (Figure 16B).

In order to confirm the specificity of the hybridization results RT-PCR experiments were performed using gene specific primers. First strand cDNA was made from each of the previously tested tissues and used as template for RT-PCR experiments run for 25 cycles. apLec-1 transcript is detectable by RT-PCR in nearly all tested tissues, but the most product was clearly visible in the spleen, followed by the intestine and lung (Figure 17). RT-PCR experiments detected apLec-2 message only in the spleen, and faint bands in the lung (Figure 17), supporting a role for this lectin in the immune system.

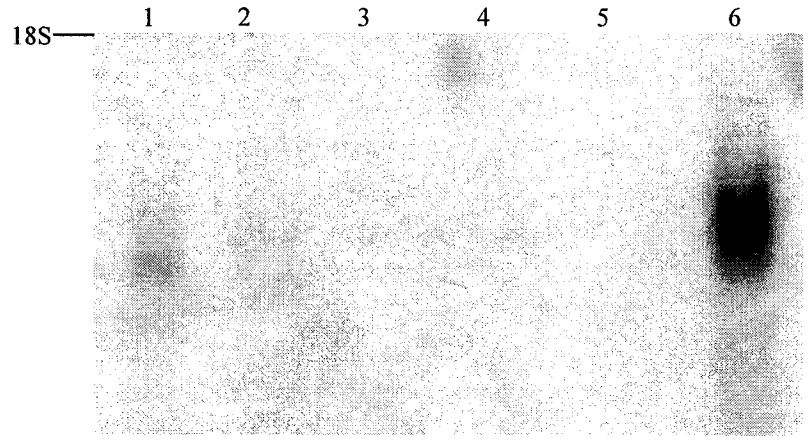
Currently there are insufficient markers and reagents to allow for the sorting of duck leukocytes into pure populations. In order to divide leukocytes into sub-populations peripheral blood mononuclear cells were isolated by Ficoll-paque density centrifugation, then separated based on their adherence to plastic. This results in an adherent population

A



apLec-1

B



apLec-2

C

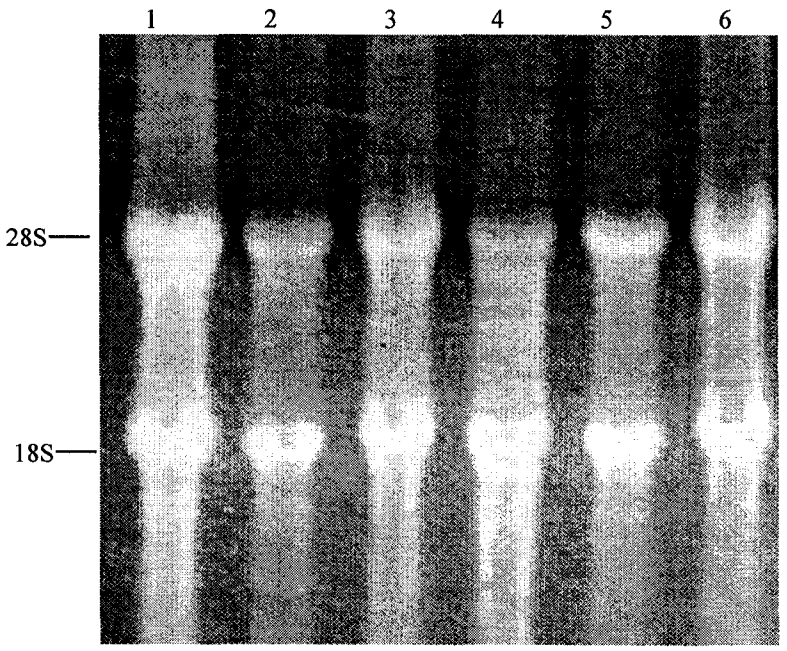


Figure 16: The apLec genes are expressed in immune and mucosal tissues. Total RNA was extracted from several tissues run on an agarose gel and transferred to a nytran membrane. Lanes correspond to 1: intestine, 2: lung, 3: heart, 4:liver, 5: kidney, 6: spleen. A: This blot was then hybridized with a probe derived from exons 1-3 of apLec-1 cDNA and exposed to film for three days. This revealed abundant apLec-1 mRNA in the spleen and reduced expression in the intestine and lung. B: The same blot was hybridized with a probe derived from exons 1-3 of apLec-2 and exposed to film for seven days. This showed apLec-2 expression mainly in the spleen. C: Image of the original agarose gel use to generate the blot above.

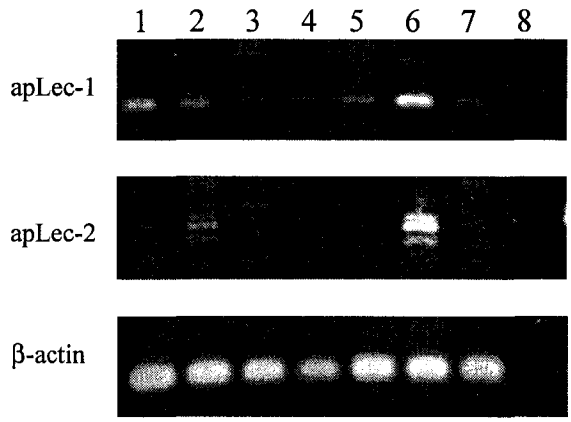


Figure 17: The apLec genes are differentially expressed. RT-PCR experiments were performed using total RNA extracted from several tissues revealing differential expression of the apLec genes. Lanes correspond to: 1: intestine, 2: lung, 3: heart, 4: liver, 5: kidney, 6: spleen, 7: smooth muscle and 8: water as template in reactions using primers specific for each apLec gene and β-actin.

enriched in monocytes and any circulating DCs and macrophages. The non-adherent population, on the other hand, should consist largely of T and B lymphocytes and possibly NK cells. In order to determine which cells express these messages total RNA was extracted from PBMCs, adherent leukocytes, non-adherent leukocytes and intestinal epithelial lymphocytes (IELS) and used as a template to make first strand cDNA.

RT-PCR experiments were then performed using primers specific for apLec-1, apLec-2 and β-actin. These RT-PCR experiments showed differences in the expression of these two lectins (Figure 18). By this method, apLec-1 message was detected in PBMCs, and in both adherent and non-adherent leukocytes but not IELs. Expression was highest in the adherent leukocytes, which was expected as this population was expected to be enriched in antigen presenting cells. Expression was also detectable in the non-adherent population, which may indicate that this message is expressed by B-cells or possibly, by contaminating monocytes and other APCs that failed to adhere. The lack of apLec-1 expression by IELs is likely to be indicative of a lack of apLec-1 expression by T cells and NK cells, as it is these cells that make up the majority of the IEL fraction in chickens (Gobel *et al.*, 2001). This also implies that APCs were not included in this IEL preparation as apLec-1 was detectable in total intestine RNA and in peripheral blood mononuclear cells but was not detected in this isolated population of intestinal lymphocytes. Under the same conditions, expression of apLec-2 message appeared highest in the adherent leukocyte population with only faint bands detected in the non-adherent population and no bands seen in IELs. Thus, like apLec-1, apLec-2 seems to be

A

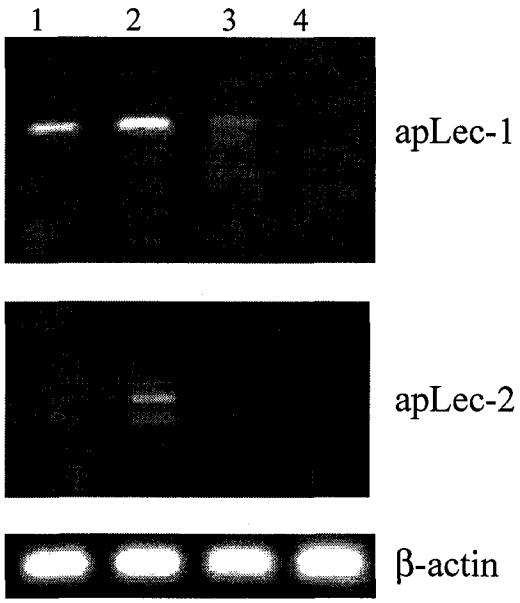


Figure 18: The apLec genes are expressed in APC containing populations. RNA was extracted from crudely sorted peripheral blood mononuclear cells (lane 1), adherent leukocytes (lane 2), non-adherent leukocytes (lane 3) and IEL (lane 4) populations and used as a template to make first strand cDNA. The cDNA was then used in PCR experiments designed to specifically amplify products spanning exons one through four of apLec-1 and exons one through five of apLec-2. β-actin was amplified from each cDNA preparation.

most abundantly expressed in the fraction enriched for APCs, though it appears to be less abundant. The virtual absence of bands in the PBMCs and non-adherent cells supports the argument that apLec-2 is a rare transcript. Dilution of the apLec-2 message in the pool of total PBMC RNA is the likely cause of the lack of strong apLec-2 bands amplified from this population. The absence of clear apLec-2 bands in the non-adherent population likely indicates that it is not expressed by T or B lymphocytes in sufficient quantities to be detected by this method and that the contribution from any contaminating APCs is not sufficient to generate a product. However, the possibility that apLec-2 expression is induced upon adherence cannot be excluded by these experiments. Like apLec-1, apLec-2 does not appear to be expressed by IELs.

To examine the level of expression upon addition of an activator of antigen presenting cells, expression of the apLec genes was examined in cells treated with lipopolysaccharide. PBMCs were isolated and incubated for 24hrs in culture medium alone or in the presence of 100 μg/mL LPS. LPS induced a considerable down-regulation of apLec-1 and apLec-2 mRNA but not the complete loss of expression (Figure 19). This downregulation is similar to results seen in studies of mammalian DCIR were various maturation or activation stimuli, including LPS, induced DCIR downregulation (Bates *et al.*, 1999) (Huang *et al.*, 2001) (Richard *et al.*, 2002). Interestingly LPS stimulation seemed to increase the proportion of the transmembrane deletion splice form of apLec-2 relative to the full-length transcript in a manner similar to the accumulation of TM-less DCIR transcripts seen in neutrophils in response to certain activating stimuli (Richard *et al.*, 2002).

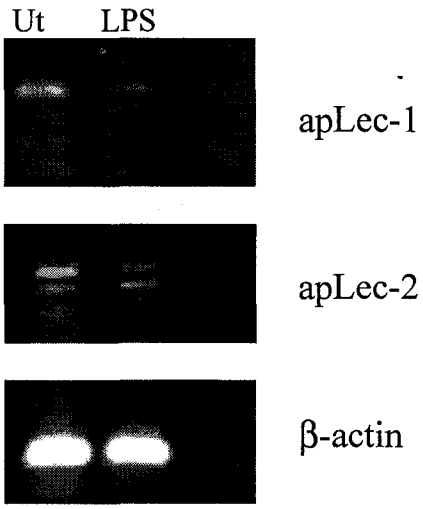


Figure 19: LPS downregulates the expression of the apLec genes. cDNA was made from PBMCs left untreated (Ut) or cultured for 24 hours in the presence of 100 μg/mL LPS and used in RT-PCR experiments to test the expression of apLec-1, apLec-2 and β-actin.

3.3 B-Lec *related genes in the duck*

The second group of lectin-encoding cDNAs identified during the sequencing of ESTs from a duck splenic cDNA library consists of five highly similar clones. These clones were designated 6D9, 14D12, 22G6, 29D1 and 33E5. The latter four clones featured identical ORFs of 621 bp, but differ in the size and sequence composition of either their 5′ untranslated region (UTR), 3′ UTR, or both (see Table 3 and Appendix 3). Clone 6D9 also featured a 621 bp ORF that shares 95% identity with the other clones. 6D9 is, however, considerably shorter than the other clones in this group due to a reduced 3′ UTR. The four longer messages bore TTTTGTA motifs, as well as AT-rich repeats in their 3′ UTRs, both of which are associated with rapid degradation of transcripts (Hamann *et al.*, 1997). This suggests that expression of these lectins is temporally regulated, at least in part, at the post-transcriptional level. The truncated 3′ UTR of clone 6D9 excludes these motifs.

Several feature of clone 6D9 suggested that this clone was incomplete. The 3′ end of this clone was considerably shorter than that of the other clones and excluded potentially important motifs. More importantly, clone 6D9 had a very unusual polyadenylation signal, in that it overlapped with the last two codons of the ORF. These two factors indicated that the brief UTR of clone 6D9 might have been an artifact. To determine if that was the case a 3′ RACE experiment was performed using a gene specific primer for clone 6D9 that differed from clone 14D12 and the others in 3 of the 5 bases at the 3′ end. Only produced products of identical size and sequence to clone 6D9 were amplified indicating that there are no expressed forms of this mRNA with a longer 3′ UTR (Figure 20).

TABLE 3: Properties of group 2 clones:

Clone designation	cDNA length (bp)	Polypeptide length (aa)	5′ UTR length (bp)	3′ UTR length (bp)	# of identical cDNAs isolated
6D9	721	207	63	37	1
14D12	1240	207	93	526	2
22G6	1148	207	93	434	1
29D1	1378	207	239	518	1
33E5	1259	207	132	506	1

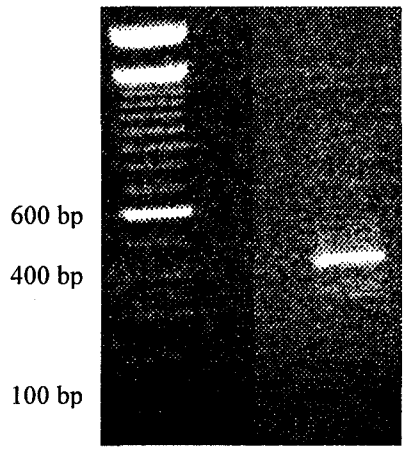


Figure 20: The truncated 3′ UTR of clone 6D9 represents the only form of this message. cDNA generated from the spleen of duck 26 was used as a template in 3′ RACE experiments. Only a single product of identical size to clone 6D9 was amplified. Sequencing of this amplified product revealed that it was identical in sequence to clone 6D9.

The 5′ end of clone 6D9 also seemed inappropriately short, as sequence in the predicted 5′ UTR of this clone was present in the ORF of the other clones. To determine if the library clone was truncated at the 5′ end RT-PCR experiments were performed using a 6D9 specific internal primer and a primer based on shared sequence in the UTR of the other clones. A larger product was obtained, however after cloning and sequencing this product two mismatches were identified. At this time it remains unclear whether these differences are an artifact, or if the RT-PCR product represents mRNA from another allele or another family member.

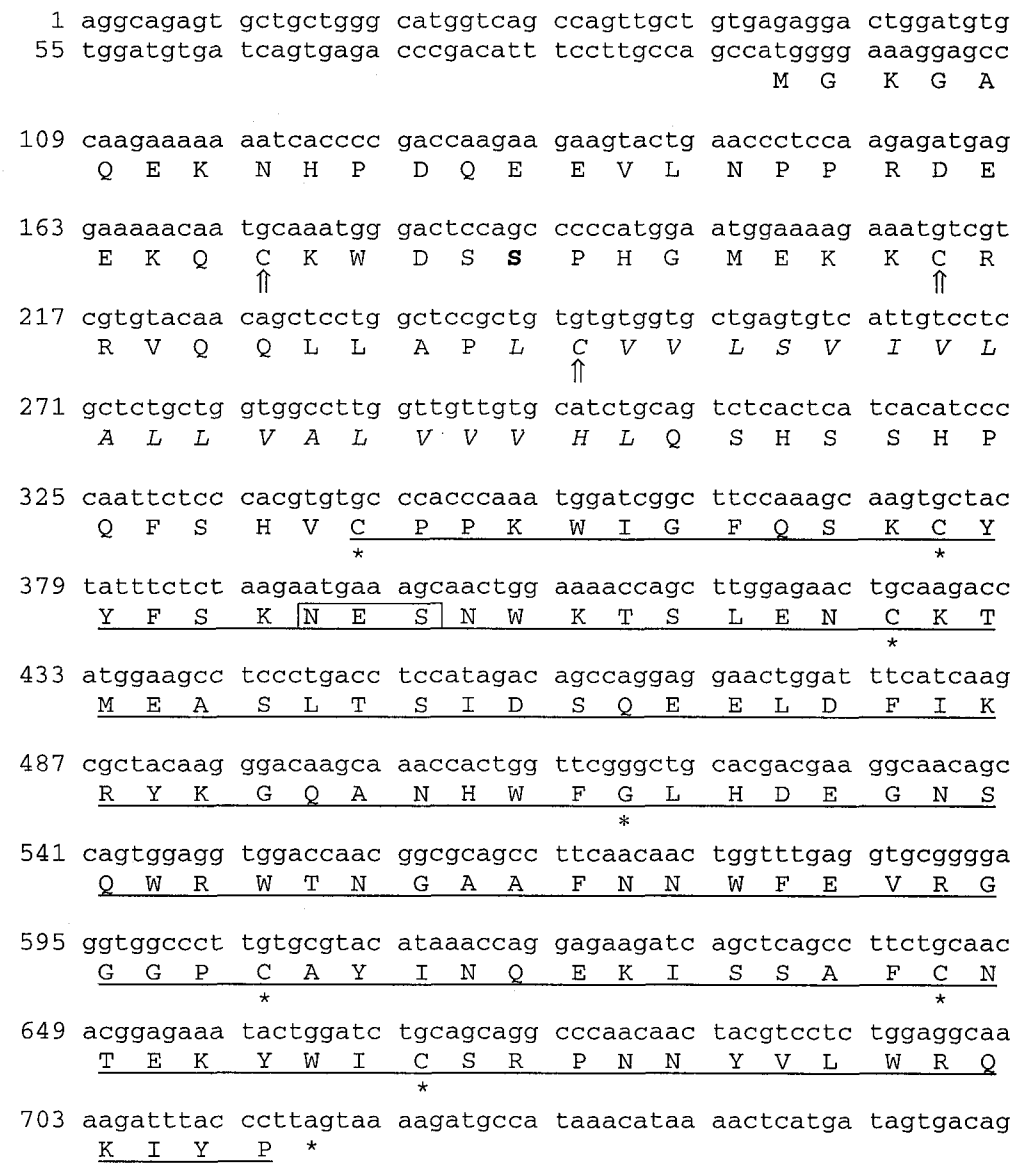
3.3.1 Predicted protein characteristics

Translation of the 6D9 ORF predicted a protein of 207 amino acids (Figure 21A) with a predicted molecular mass of 24.00 kDa; the others, represented by clone 14D12, encoded a 207 amino acid protein (Figure 21B) with a predicted molecular mass of 23.99 kDa before post-translational modifications. The proteins encoded by 6D9 and 14D12, though distinct, showed a high degree of sequence identity and shared several structural features. Both proteins contain a hydrophobic region that is predicted to be a transmembrane domain, but lacked an N terminal signal peptide, thus, these proteins are likely to be transmembrane proteins of a type II orientation. The polypeptides encoded by 6D9 and 14D12 both contained the six conserved cysteine residues and conserved glycine typical of C-type lectin long-form carbohydrate recognition domains in their extracellular portion. The conserved glycine was in the context of a variant "WIGL" motif of the sequence WFGL in both of these proteins. This variant is in keeping with the aromatic residue – hydrophobic residue – glycine – hydrophobic residue consensus that forms this conserved β strand. A twelve amino acid neck region separated the CRD

A

```
  1 gttgctgtg agaggactg gatgtgtgg atgtgatca gtgagaccc gacatttcc
 55 ttgccagcc atggggaaa ggagcccaa gaaaaaaat caccctgac caagaagaa
              M  G  K   G  A  Q   E  K  N   H  P  D   Q  E  E

109 gtaatgaac cttccaaga gatgaagag aaacaatgc aaatggggc ttcagcccc
     V  M  N   L  P  R   D  E  E   K  Q  C   K  W  G   F  S  P
                                          ⇑
163 catgggatg aaaaagaaa tgtcgtcgt gtaaaaaag ctcctgact ccgctgtgt
     H  G  M   K  K  K   C  R  R   V  K  K   L  L  T   P  L  C
                         ⇑                                   ⇑
217 gtggtgctg actgtcctt gtcctcgct ctggtggtg gccttggtt gttgtgctt
     V  V  L   T  V  L   V  L  A   L  V  V   A  L  V   V  V  L

271 ctgcagtct cactcatca catccccaa ttctccgac gtgtgccca gacaaatgg
     L  Q  S   H  S  S   H  P  Q   F  S  D   V  C  P   D  K  W
                                              *
325 atcggcttc caaagcaag tgctactat ttctcggag gatgaaagc aactggaaa
     I  G  F   Q  S  K   C  Y  Y   F  S  E   D  E  S   N  W  K
                         *
379 accagcttg gagaactgc aaggccatg gaagcctcc ctgacctcc atagacagc
     T  S  L   E  N  C   K  A  M   E  A  S   L  T  S   I  D  S
                      *
433 caggaggaa ctggctttc atcaagcgc ttcaagggc caagcaaac cactggttc
     Q  E  E   L  A  F   I  K  R   F  K  G   Q  A  N   H  W  F

487 gggctgcac gacgaagac aacagccag tggaggtgg accaacggc gcagccttc
     G  L  H   D  E  D   N  S  Q   W  R  W   T  N  G   A  A  F
     *
541 aacaactgg tttgaggtg cggggaggt ggcccttgt gcgtacata aaccaggag
     N  N  W   F  E  V   R  G  G   G  P  C   A  Y  I   N  Q  E
                                          *
595 atcagctca gccttctgc aacacggag aaatactgg atctgcagc aggcccaac
     I  S  S   A  F  C   N  T  E   K  Y  W   I  C  S   R  P  N
                      *                       *
649 aactacgtc ctctggagg caaaagatt taccccgaa taaagatct tataaacat
     N  Y  V   L  W  R   Q  K  I   Y  P  E   *

703 aaaaaaaaa aaaaaaaaa a
```

B



C

```
   6D9 MGKGAQEKNHPDQEEVMNLPRDEEKQCKWGFSPHGMKKKCRRVKKLLTPLCVVLTV 56
14-D12 ..............L.P.........DS.....E......QQ..A......S. 56

   6D9 LVLALVVALVVVLLQSHSSHPQFSDVCPDKWIGFQSKCYYFSEDESNWKTSLENCK 112
14-D12 I....L......H..........H...P............KN............ 112

   6D9 AMEASLTSIDSQEELAFIKRFKGQANHWFGLHDEDNSQWRWTNGAAFNNWFEVRGG 168
14-D12 T.............D.................G..................... 168

   6D9 GPCAYINQE-ISSAFCNTEKYWICSRPNNYVLWRQKIYPE                 207
14-D12 .........K.............................                 207
```

Figure 21: Sequence and alignment of clones 6D9 and 14D12. A: cDNA sequence of clone 6D9. Residues in bold represent the predicted sequence of the missing 5′ segment of this clone based on alignments with the RT-PCR products and other clones from this family. Predicted translation is indicated beneath the ORF. B: cDNA and predicted amino acid sequence of clone 14D12. In both A and B the transmembrane region is indicated by italics. The CRD is underlined. The six conserved cysteines typical of long-form CRDs are indicated with asterisks. The conserved glycine is indicated by a *. The serine predicted to be phosphorylated in the cytoplasmic domain of these proteins is indicated in bold. Potential palmitoylation sites are indicated by ⇑. The predicted N-linked glycosylation site is boxed. Polyadenylation signal is double underlined. Rapid degradation motifs are in italics. C: Amino acid sequence alignment of 6D9 and 14D12. Dots indicate sequence identity.

from the transmembrane domain. These properties are typical of the immunoreceptors encoded in the NKC in mouse and man. Like many NKC encoded proteins these duck proteins lacked the residues involved in calcium ligation that characterize classical lectins, which may mean that they recognize non-carbohydrate ligands, or bind carbohydrates without dependence on calcium.

6D9 featured a cytoplasmic domain of forty-nine amino acids, but lacked ITIM or ITAM signaling motifs. There is a net positive charge of five within the cytoplasmic domain, but there are no positively charged residues within the transmembrane domain to mediate association with adaptor molecules such as DAP-12. There is, however, a serine residue in the context SPH that may be phosphorylated, and thus may have a role in the function of this protein. 14D12 also has a cytoplasmic domain of forty-nine amino acids. Like 6D9, the 14D12-encoded protein does not bear any consensus immunoreceptor signaling motif. However, there is a histidine residue at the C-terminal end of the predicted transmembrane domain that may act to recruit adaptor molecules, though this is unlikely as the docking site for adaptor molecules in other immunoreceptors is in the central or N-terminal portion of the TM and the histidine in this position is more likely to interact with the phosphate groups of the phospholipids in the outer leaflet of the plasma membrane. This residue is not present in the translation of 6D9. 14D12 has a lower net charge on its cytoplasmic domain, with eleven positively charged residues balanced by nine negatively charged ones to give a net positive charge of two to this domain. The 14D12-encoded protein shared the potentially phosphorylated serine that is present in 6D9.

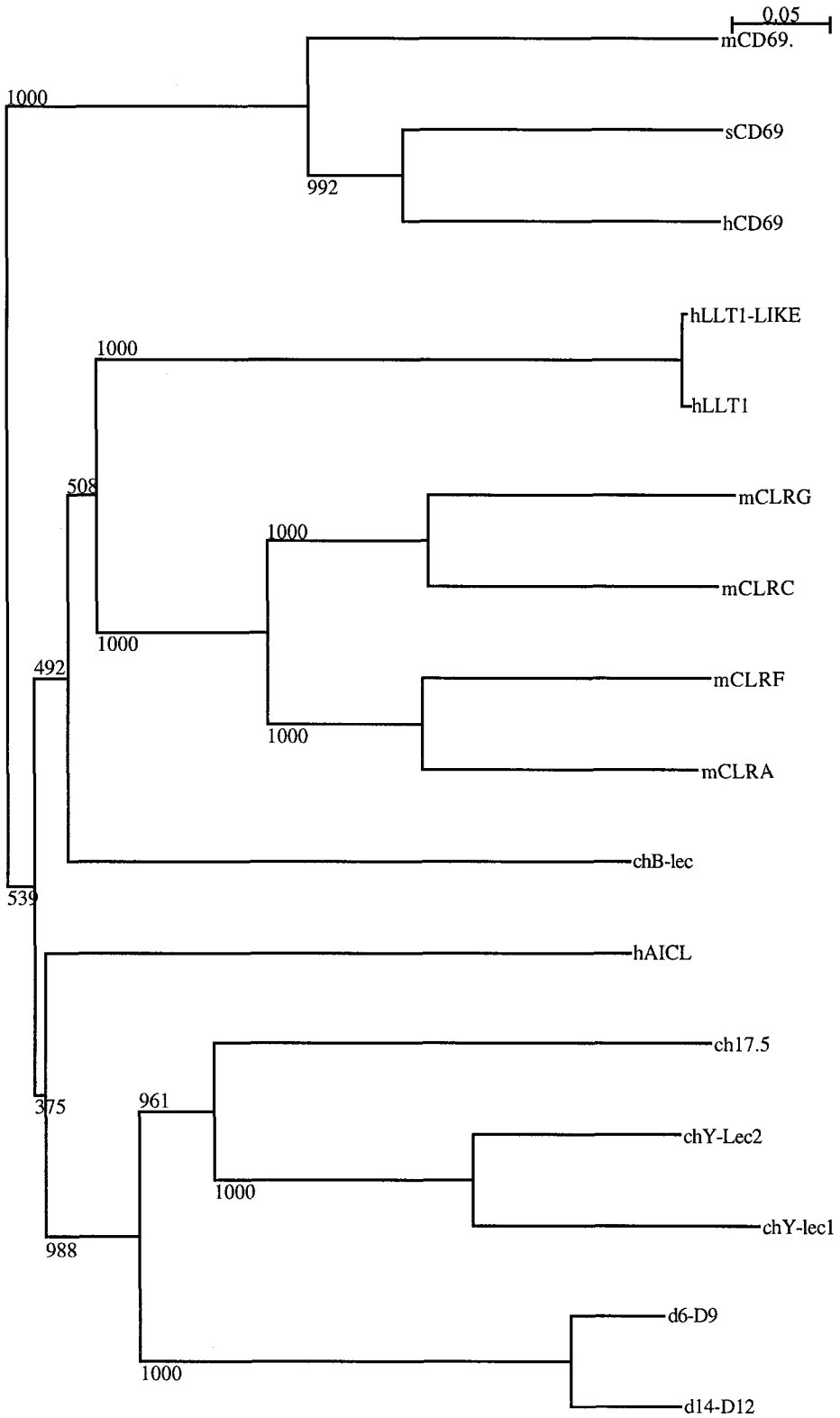
Both of the predicted proteins share two cysteine residues at equivalent positions within the cytoplasmic tail, sufficiently near the transmembrane domain to be palmitoylated, as well as an additional cysteine just inside the transmembrane region that may also be palmitoylated. This modification is likely to influence the localization of these proteins to detergent-resistant microdomains within the plasma membrane (Resh, 1999). Of the two proteins, only the one represented by 14D12 undergoes N-linked glycosylation at an NES site within the CRD.

A translated BLAST search with these sequences revealed strong similarity to the product of the chicken gene 17.5 (Bernot *et al.*, 1994), located within the *Rfp-Y* region, as well as the recently described YLec2 and to a lesser extent YLec1 (Rogers *et al.*, 2003), and the chicken B-Lec gene product, encoded in the *Mhc B* region in chicken (Kaufman *et al.*, 1999). This analysis also revealed similarity to AICL, LLT1, members of the CLR family, CD69 and several other related mammalian genes.

Clustal X analysis was performed to determine the relationship between these novel duck receptors, related avian lectins, and the mammalian NKC encoded immunoreceptors. A dendrogram revealed that these lectins group closely with human AICL and associated lectins such as the rodent CLR family of receptors and human LLT1 and are more distantly related to the activation antigen CD69 (Figure 22).

Amino acid alignment of these lectins revealed shared characteristics and differences between members of this group (Figure 23). The CRDs of 6D9 and 14D12 showed the most homology with Y-Lec-2 (77%), gene 17.5 (72%), B-Lec (64%), AICL (65%), CLRA (62%), and LLT1 (62%), however unlike CLRA and LT1 these duck

A



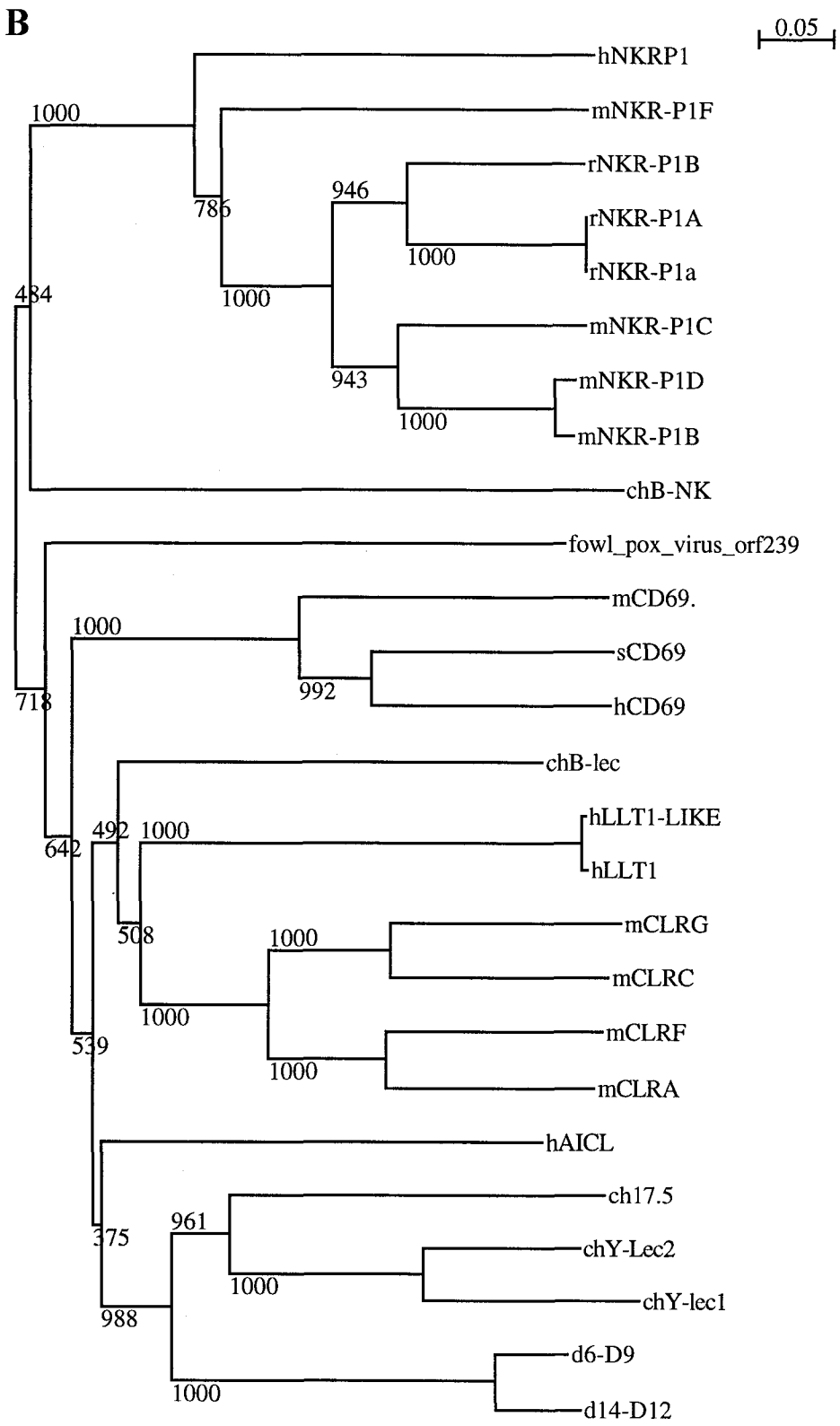


Figure 22: Dendrograms showing the relationship between duck lectins 6-D9, 14-D12 and immunologically relevant lectins found in the NKC of mouse and man. Both dendrograms are subtrees of larger analyses including most lectin-like receptors. A: A tree generated using CRD sequence alone groups avian lectins with AICL and the LLT1/CLR receptors. B: Analysis based on entire amino acid sequence shows these receptors are also related to mammalian NK receptors, albeit more distantly.

```
      6D9                                                                MGKGAQE  7
    14D12                                                                .......  7
gene 17.5 MPLFLSFSPRSLREVLAKKSAPPAPLCPQPDPSLLLSLHAAGAVPHLYDATEEKER 56
   Y-Lec1                                                                ..E.D.Q  7
     CLRB                                                                   MCVT  4
     CLRG                                                                      M  1
     CLRC                                                                      M  1
     mCD69                                                               .DSENC  6
     hCD69                                                               .SSENC  6

      6D9 KNHPDQEEVMNLPRDEEKQCKWGFSPHGMKKKCRRVKKLL-TPLCVVLTVLVLALV  62
    14D12 .........L.P..........DS.....E......QQ..-A......S.I....L  62
gene 17.5 LSPSPPR.ATTREG...R.SQR.SGCSELRQNR...--.C-VA.SA.PCM......  109
   Y-Lec2 MPPVGEGDQQETFSEHQAATEPRSWC.GAGRRRS..--Q.-IAV.AA.GA.I.V..  53
    B-Lec                                                  MAA.FTVL  8
     CLRA  MLKTDLTAPDCLQEG.MGK.LQAKCL.IISAASC.KLYCCYGVIM....A.S..S  55
     AICL                                                MMTKHKKCFII.  12
   Y-Lec1 ETFSEPKAALQPLGQSGGH-Q..SWC..TGRRRS..--Q.-IAA.AA.GT.S.V..  59
     LLT1   MH.SNN.EKDITPS.LPANP.CLHSKEHSIKATLIWR.-FF.IMF..II.CGM.  53
     CLRF MLQRADLTAADCLQEG.MGK.IQGKCFRIISTVSP..LYCCYGVIM....A.I..S  56
     CLRB .ASLPMLSPTGS.QEV.VGKILQGKR..TISPESCA.LYCYYGVIM....A.I..S  60
     CLRG NAQCV.KPEEGNGPLGTGGKIVQGKCFRIISTVSP..LYCCYGVIM....A.I..S  57
     CLRC NAQCLKKPEEGESSPGTGDKILQRNSLRAISPESSA.LYCCCGVIM....A.V..S  57
     mCD69 SITENSSSHLERGQKDHGTSIHFEKH.EGSIQVSIPWAV.IVV.ITS.IIALI..N  62
     hCD69 FVAENSSLHPESGQENDATSPHFSTR.EGSFQVPVLCAVMNVVFITI.IIALI..S  62

      6D9 VALVV-VLLQSHS-SHPQFSDVCPDKWIGFQSKCYYFSEDESNWKTSLENCKAMEA 116
    14D12 .....-.H.....-......H...P.............KN............T... 116
gene 17.5 AV-I.-LQRP.C.-PR.P..H...NA.V...G......DT..D.NS.R.H.HRLG. 162
   Y-Lec2 .IST.-CRQVPVP-PF.D.AHA..NA.V...G......KE.ND.NS.R.H.N.HG. 107
    B-Lec LITA.AFAV.AFQ-P...PCAQ..FD....RG............TS.QN..SALG. 63
     CLRA .T.S.RKKKPVME-.CEPCYA..SSG....GN..F.....MG..TF.QSS.I.LD. 110
     AICL GV.ITTNIITLIVKLTRDSQSL..YD.....N......KE.GD.NS.KY..STQH. 68
   Y-Lec1 .IST-------------D.AHA..NA.V...G....L.KE.ND.NS.R.H.N.HG. 102
     LLT1 A..SAIRANCHQE-PSVCLQAA..ES.....R..F...D.TK..TS.QRF.DSQD. 108
     CLRF ...S.RNKIPAME-DREPCYTA..SG....G...F.....MG..TF.QSS.V.SNS 111
     CLRB ...SATKTE.IPV-NK--TYAA..QN...VEN..F....YP...TFAQAF.M.Q.. 113
     CLRG ...STKKTE.III-NK--TYAA.SKN.T.VGN..F...GYPR..TFAQAF.M.Q.. 110
     CLRC ...PATKTE.ILI---NKTYAA..KN...VGN..F....YT...TFAQTF.M.Q.. 110
     mCD69 .GKYNCPG.YEKLE.SDHHVAT.KNE..SYKRT..F..TTTKS.ALAQRS.SEDA. 118
     hCD69 .GQYNCPGQYTF.MPSDSHVSS.SED.V.Y.R...FI.TVKRS.TSAQNA.SEHG. 118

      6D9 SLTSIDSQEELAFIKRFKGQANHWFGLH-DEDNSQWRWTNGAAFNNW---FEVRGG 168
    14D12 ..........D....Y..........-..G..............---..... 168
gene 17.5 ..ATLDTK..ME.MLQYQRPADR.I...RA.GDEH.T.AD.S..T.R-PV..L... 217
   Y-Lec2 ..AT.G.A..MD.MM..Q.P..C.I...RE.EDA..T.SD.T..T..---..L... 160
    B-Lec ..AVF..A.D.S.TM.H..SSP..V..SREGKEHP.E.V.RSPLSH---L.Q.Q.D 116
     CLRA H.ALF..L...N.L..Y..ASD..I...RESSEHP.I..DNTEY..---LVLT... 163
     AICL D..I.DNI..MN.LR.Y.CSSD..I..-KMAK.RTGQ.VH..T.TK---S.GM..S 120
   Y-Lec1 ..AT.G.A..MD.MM..Q.P..C.I...WE.EDAL.T.S.VT..T..RGSPSFSQC 158
     LLT1 D.AQVE.FQ..N.LL.Y..PSD..I..SREQGQ-P.K.I..TE---.TRQ.PIL.A 160
     CLRF H.ALFH.L...N.L..Y..TSD..I...RASTQHP.I..DNTEYS.---LVLT... 164
     CLRB Q.ARF.N.D..N.LM.YKANFDS.I...RESSEHP.K..DNTEY..---TIPI..E 166
     CLRG Q.ARF.NE...I.L.....DFDC.I...RESSEHP.K...NTEY..---MNPIL.V 163
     CLRC Q.ARF.NEK..N.L..-HMNSS..I...R.SSEHP.R..DNTEY..---T.LIQ.D 162
     mCD69 T.AV...EKDMT.L..YS.ELE..I..-KN.A.QT.K.A..KE..S.---.NLT.S 170
     hCD69 T.AV...EKDMN.L..YA.REE..V..KKEPG-HP.K.S..KE....---.N.T.S 170
```

```
      6D9 GPCAYINQE-ISSAFCNTEKYWICSRPNNYVLWRQKIYPE                207
    14D12 .........K............................                 207
gene 17.5 .R...L.GDG....L.HS..F.V...ADS..R..KGTN.Q                257
   Y-Lec2 .R...L.GDR...SL.HLH.H.V...ADH....K..VH.Q                200
    B-Lec .L...LGDAGL..SH.S.RRN.V.TK.ALQKPRKNFCIST                156
     CLRA .E...LSNRG.YNSSGDIH.K...NK....T.Q.PL.VNPG                204
     AICL EG...LSDDGAAT.R.Y..RK...RKRIH                            149
   Y-Lec1 IGA.FPFSSLLGTGLSCEVEADVRT                                183
     LLT1 .E...L.DKGA...RHY..RK....KSDIH.                          191
     CLRF .E.GFLSDNG...GRSY.HRK....KFVSSCKS.VGSV.RH                205
     CLRB ERF..L.NNG...TRIYSLRM....KL.S.S.HC.TPFFPS                207
     CLRG .RY..LSSDR...SRSYINRM....KL...N.HC.TPPV                  202
     CLRC .E.GFLSDNG...SRDYIPRK.....SS..M.QC                       196
    mCD69 .R.VSV.HKNVTAVD.EANFH.V..K.SR                            199
    hCD69 DK.VFLKNTEV..ME.EKNL....NK.YK                            199
```

Figure 23: Amino acid alignment comparing the duck AICL-like lectins, chicken lectins and mammalian CD69 related lectins.

clones retained all six conserved cysteines. The lowest identities in the CRD were with murine CD69 (~33%) and Y-Lec-1 (36%). However, when comparing the cytoplasmic domains CD69, CLRA, LLT1 and AICL shared the most identity while the chicken Y-Lec genes were only 6% and 3% identical respectively. In light of the overall similarity to AICL and the results of the phylogenetic analysis, clone 6D9 was named AICL-like1 and 14D12 called AICL-like2.

In fact comparison of these amino acid sequences and predictions of their secondary structure to those of lectin-like receptors with known crystal structures suggested that the CRD of these receptors fold into a similar tertiary structure to that of CD69 (Figure 24), which may indicate a homologous function for these receptors.

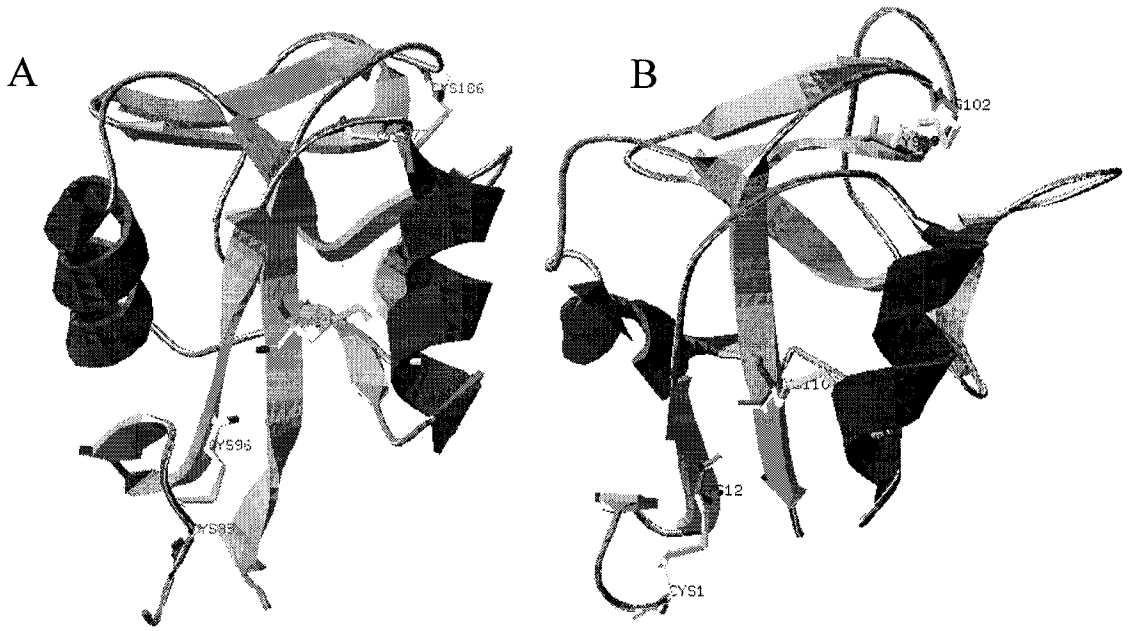


Figure 24: The CRDs of the duck AICL-like receptors resemble that of CD69. **A:** Crystal structure of CD69 compared to **B** the predicted crystal structure for AICL-like2. This analysis reveals extensive similarity between the structures of these receptors despite the divergence of their primary sequences.

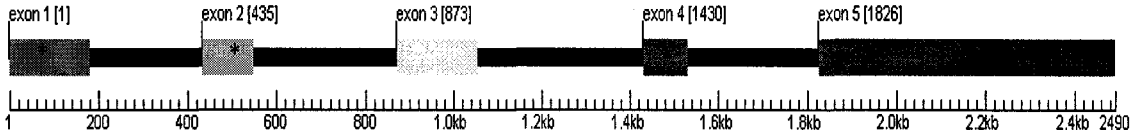
3.3.2 Genomic analysis

To obtain the genomic sequence for the AICL-like genes PCR reactions were performed using genomic DNA from ducks 64, 95, 105, 129 and 132 using a forward primer that begins at the 28th nucleotide of the open reading frame of both AICL-like genes and a reverse primer based on a region of conserved sequence that ends 48 nucleotides from the stop codon of AICL-like2. The PCR generated a major product of approximately 1.8kb in these ducks. These products were cloned into the Topo 2.1 vector (InVitrogen). Multiple clones from each individual were selected and sequenced.

Analysis of the coding region at the genomic level confirmed that these receptors are encoded in five exons (Figure 25A). The first exon encoded most of the cytoplasmic domain, exon two encoded the rest of the cytoplasmic domain and the majority of the transmembrane domain, exon three encoded the final residues of the transmembrane domain, the entire neck region and part of the CRD, while exons four and five encoded the remainder of the CRD. This genomic organization further supports the relationship between the duck genes and the AICL, CLR and CD69 genes, which also consist of five translated exons and have similarly short neck domains that are encoded by the same exon as the N-terminal portion of the CRD (Santis *et al.*, 1994). This was in contrast to many other NKC encoded receptors such as the NKR-P1 and the LY49 genes which are divided into six coding exons with the third exon encoding only the neck region and exons four through six encoding the CRD. The intron/exon junctions are shown (Figure 25B).

The genomic sequences revealed polymorphism between individuals. The degree of differences between individuals is relatively low (Figure 25C), however most of the

A



B

	5′ splice site	3′ splice site
Intron 1	GGG/GTAAT	TCTCTTCTACACCAG/A
Intron 2	TG/GTGAGT	TTCTCCTTTTACAG/TT
Intron 3	CTG/GTGAG	TTTTAGCGATAAAG/GA
Intron 4	AACTG/GTC	GTGTGTCTGCATwG/CA

C

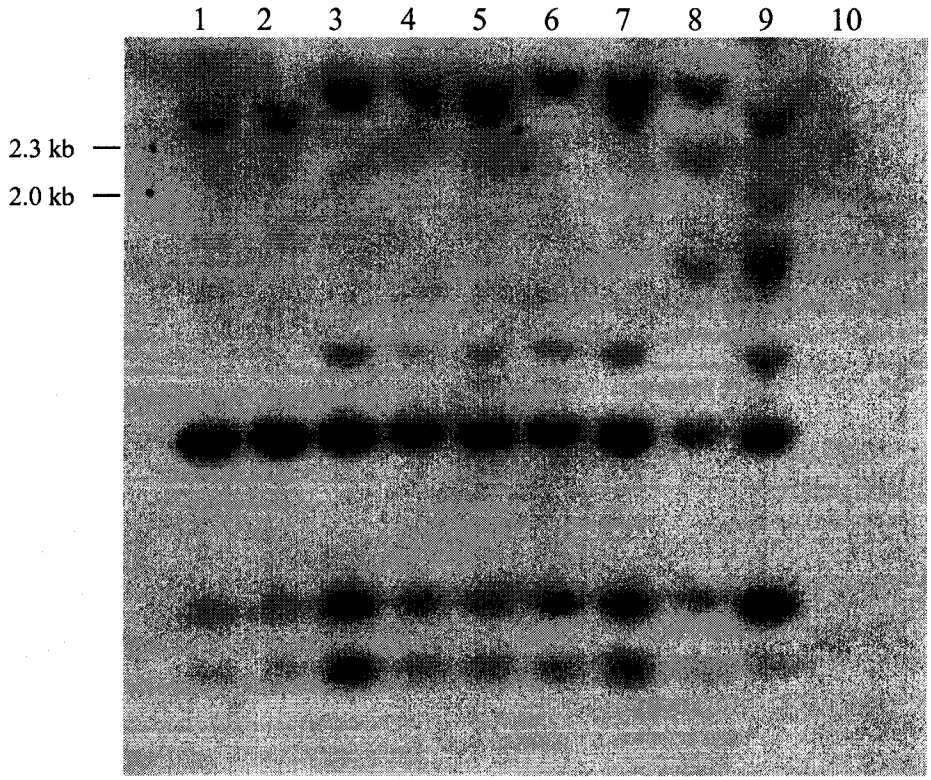
	105	129-2	132-1	132-2	95-1	95-2	95 (6D9)
105		94%	94%	90%	94%	94%	96%
129-2	94%		93%	85%	96%	92%	95%
132-1	94%	93%		89%	92%	95%	93%
132-2	90%	85%	89%		82%	86%	86%
95-1	94%	96%	92%	82%		91%	99.3%
95-2	94%	92%	95%	87%	91%		92%
95 (6D9)	96%	95%	93%	86%	99.3%	92%	

Figure 25: The AICL-like genes are encoded in five exons and are polymorphic. A: Schematic of the genomic organization of the AICL-like genes. Asterisks indicate sites of non-silent polymorphisms detected between individuals. B: Intron exon junctions of AICL-like 2 based on the sequence from 95-1 and 95 (6D9). C: Pairwise identity matrix of sequenced AICL-like alleles based on nucleotide sequences.

polymorphisms are non-silent implying that these genes are under selective pressure. The results from duck 95 suggest that AICL-like 1 and AICL-like 2 may in fact be expressed from separate loci. Duck 95 yielded three homologous AICL-like products. Two of these products represented by clone 95-1 and the sequence designated 95 (6D9) appear to be allelic forms of AICL-like 1 and clone 95-2 resembles AICL-like 2, indicating the presence of at least two closely related loci.

Southern analysis supported the existence of additional loci. Genomic DNA from several individuals was digested with a single restriction enzyme, transferred and hybridized with the 14D12 insert. This revealed restriction fragment length polymorphism between individuals, supporting the finding that these genes are polymorphic (Figure 26A). Genomic DNA from a single individual was digested with a panel of restriction enzymes and transferred to a Biodyne® B membrane. The membrane was then hybridized with a cDNA-derived probe that excluded the CRD encoding portion of the message. This revealed a simple pattern of only a few hybridizing bands under low stringency conditions, however there are more bands present than can be accounted for based on the restriction sites in the known genomic sequence (Figure 26B). This indicates that there is likely to be more than one family member in the duck genome.

A



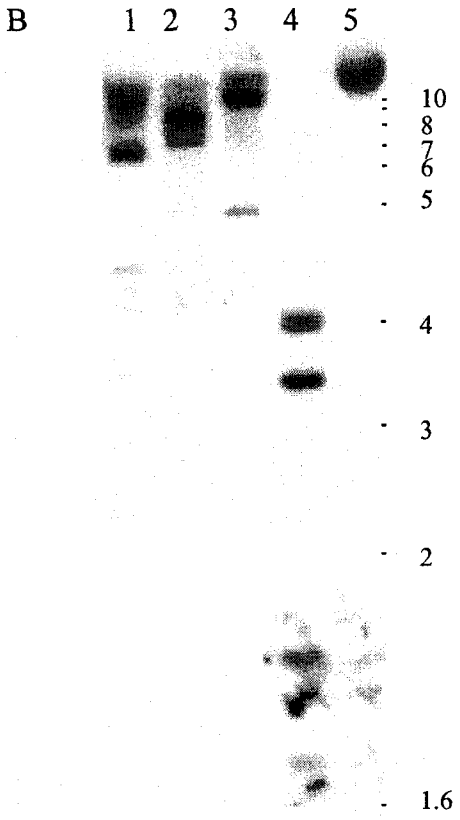


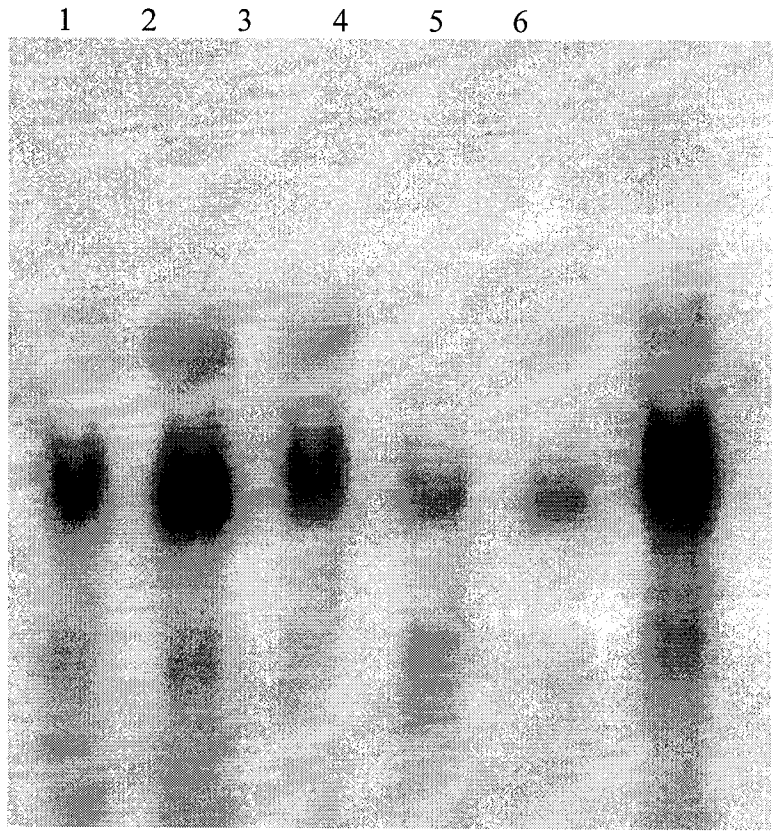
Figure 26: The duck AICL-like genes are encoded by multiple loci and are polymorphic. **A** Genomic DNA from nine ducks (lanes 1-9) and from one chicken (lane 10) was completely digested with Pst 1 and run on a 0.8% agarose gel in TBE. The DNA was transferred to a Nytran membrane and hybridized with a radiolabeled probe made from the CRD encoding portion of 14-D12 cDNA. Hybridization revealed restriction fragment length polymorphisms between individuals. **B** Genomic DNA from a single duck was digested with various restriction enzymes and transferred to a Biodyne¤ B membrane. Lanes correspond to 1: Xho I, 2: Xba I, 3: Spe I, 4: Pvu II and 5: Not I. This blot was hybridized with a probe derived from exons 1-3 of 14D12. There are more hybridizing bands shown than can be explained by the known restriction enzyme cut sites within these genes, thus it is likely that clone 14D12 represents a small gene family.

3.3.3 Expression of AICL-like messages

Total RNA was extracted from a panel of tissues from a single 6 week old duck congenitally infected with duck Hepatitis Virus and examined for expression of these transcripts by Northern blotting and reverse transcription PCR. Northern analysis indicated that related transcripts are present in most tissues (Figure 27). Which, in light of the likelihood that these messages are rapidly degraded and the degree of similarity to mammalian activation antigens, implies either a relatively large proportion of activated infiltrating lymphocytes in these tissues or that these genes, like the CLR family may encode self-ligands recognized by NK cells and therefore be expressed by a wide variety of cell types. Another possibility is that this expression pattern seen is due to the fact that the tested individual was congenitally infected with Hepatitis B virus. Although tolerance to HBV is expected in a congenitally infected duck, the expression of the AICL-like messages seen here may be the result of this infection and not representative of the normal state.

Because these two messages are too similar to differentiate by hybridization RT-PCR experiments were performed to study the expression of each of the individual AICL-like genes (Figure 28). First strand cDNA was made from RNA obtained from the intestine, lung, heart, liver kidney and spleen of the same duck used for the northern blot. RT-PCR was then carried out for 25 cycles with primers specific for AICL-like 1, AICL-like 2 or β actin. Both AICL-like messages are expressed to varying extents in all of the tissues tested. AICL-like 1 messages showed more variability between tissues with the greatest amount of product produced in the spleen, kidney and lung.

A



B

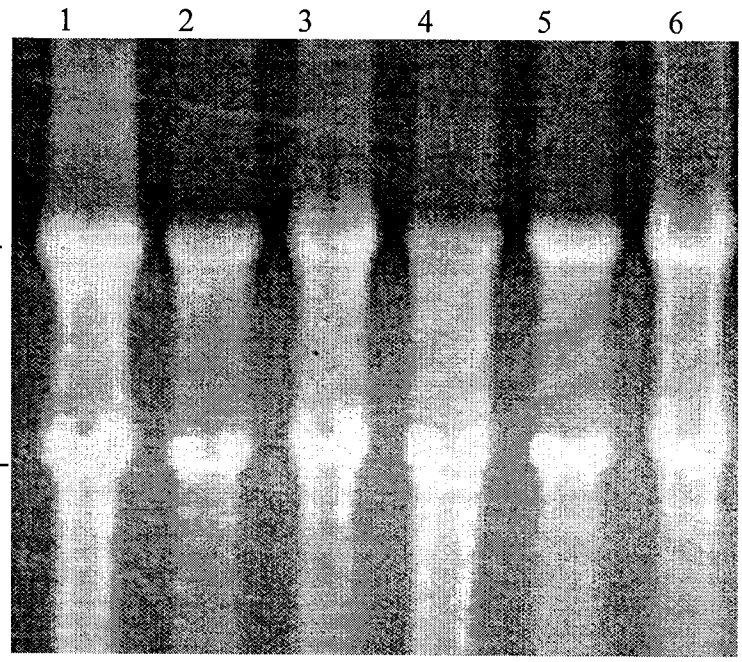


Figure 27: Duck AICL-like messages are expressed in most tissues in a duck congenitally infected with duck hepatitis B virus, with particularly strong expression in the spleen and in mucosal tissues. Total RNA was extracted from several tissues. Lanes correspond to 1: Intestine, 2: lung, 3: heart, 4: liver, 5: kidney, and 6: spleen. Northern blot hybridized with a probe derived from 14D12 exons 1-3. As this probe is unlikely to distinguish between AICL-like 1 and AICL-like 2, hybridizing bands likely correspond to both messages, with the smallest band likely representing 6D9 message.

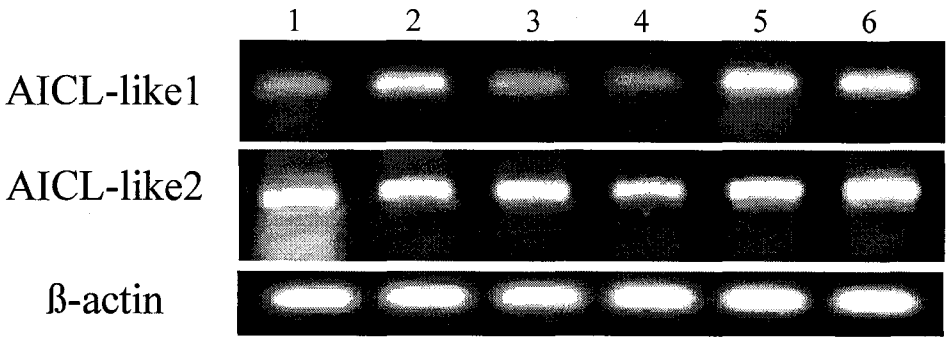


Figure 28: Both AICL-like messages are broadly expressed. Total RNA was extracted from several tissues and used as a template to make first strand cDNA. Lanes correspond to cDNA derived from 1: intestine, 2: lung, 3: heart, 4: heart, 5: kidney, and 6: spleen. This cDNA was used as template in 25 cycle RT-PCR experiments with primers specific for AICL-like 1, AICL-like 2 or β-actin. This analysis confirms that these lectins are expressed in several tissues and shows that this expression pattern extends to both AICL-like genes.

The expression of AICL-like 1 and AICL-like 2 was also examined in crudely sorted leukocyte populations. RT-PCR was performed using AICL-like 1 specific primers on cDNA generated from peripheral blood mononuclear cells, adherent and non-adherent leukocyte populations and either untreated or concanavalin A treated intestinal epithelial lymphocytes. Products of varying intensities were detected in all populations (Figure 29). This revealed that there are only slight differences in expression between the AICL-like 1 and AICL-like 2 messages, likely due to the lack of the rapid degradation motifs in AICL-like 1 that are present in AICL-like 2

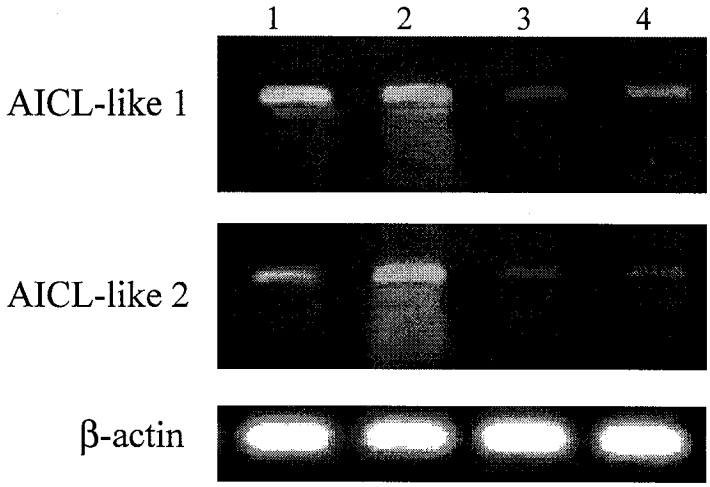


Figure 29: AICL-like genes are expressed in a variety of leukocyte populations. First strand cDNA was made from total RNA extracted from different leukocyte populations. RT-PCR experiments were then performed using gene specific primers for AICL-like 1, AICL-like 2, and β actin. Populations tested include adherent leukocytes (lane 1), non-adherent leukocytes (lane 2), Intestinal epithelial lymphocytes (lane 3) and IELs treated with concanavalin A for one hour (lane 4). Con A treatment appears to upregulate AICL-like expression slightly.

4. Discussion:

4.0. Lectin-like immunoreceptors cloned through the EST project

During the course of an EST project examining genes expressed in the spleen of a White Pekin duck two related groups of genes were discovered. The products of these genes belong to the C-type lectin superfamily.

4.1.1 The apLec genes represent homologs of APC genes encoded near the NKC

One group is most similar to the type II lectin-like immunoreceptors expressed on antigen presenting cells. One member of this group is strikingly similar to DCIR; the only ITIM bearing APC associated receptor identified to date. DCIR and apLec-1 both bear a cytoplasmic ITIM of identical sequence. However, the carbohydrate recognition domain of apLec-1 is more similar to human CLECSF8/ MCL (Arce *et al.*, 2004), murine MCL (Balch *et al.*, 1998) and Blood Dendritic Cell Antigen 2 (BDCA) (Dzionek *et al.*, 2001) than to that of DCIR. Despite the relative divergence of the CRDs of these receptors, they do retain several important common features. The CRDs of human DCIR, murine DCIR and apLec-1 shared an N-linked glycosylation site. apLec-1 contained six of the nine residues originally identified as being unique to the CRD of DCIR (Bates *et al.*, 1999), many of which are absent from related receptors such as MCL and Mincle. Like DCIR, the CRD of apLec-1 contains all six residues involved in calcium binding site 2 and the same EPS motif in the specificity-determining region of the CRD; these features are not shared by NKC encoded, ITIM bearing lectins.

Preliminary results indicate that, as is the case for DCIR, apLec-1 message is downregulated by DC maturation stimuli, or at the very least LPS. Isolated PBMCs were cultured in the presence of LPS at a concentration of 100μg/mL for 24 hours and then the

RNA was extracted from these cells. When RT-PCR experiments were performed the untreated cells showed greater expression of apLec-1 message than did the LPS treated cells despite similar levels of β actin message. These experiments should be repeated in a more refined culture system to confirm these results.

In light of the conservation of such potentially functionally important features, the differences between apLec-1 and DCIR indicate either a great amount of divergence between the DCIR genes in birds and mammals or apLec-1 represents a novel ITIM bearing APC-associated lectin. Given that the DCIR, DCAR, MCL and Mincle genes in mammals are presumed to have arisen through the duplication and divergence of a common ancestral gene it is relatively unsurprising that the avian form of this inhibitory receptor resembles all of these mammalian genes. Nonetheless, the conservation of an inhibitory receptor with the characteristics of a DC associated lectin between ducks and mammals implies that it performs an important function on APCs.

A second clone encoded a related receptor named apLec-2. The CRD of apLec-2 shows a general similarity to the CRDs of APC lectins like DCIR, DCAR, DLEC, MCL and Mincle and is highly similar to that of apLec-1. In fact, the degree of similarity between the CRDs of these two proteins is reminiscent of the similarity between activating and inhibitory pairs of receptors like Ly49H and Ly49I (90% identity) (Arase *et al.*, 2002) or DCAR (Kanazawa *et al.*, 2003) and DCIR (89% identity) (Kanazawa *et al.*, 2002). The probable relationship between apLec-1 and 2 is supported by the presence of a positively charged residue in the transmembrane domain of apLec-2. This is a feature common to activating immunoreceptors; this positively charged residue in the receptor mediates association with ITAM-bearing adaptor molecules, which bear a

negative charge within their transmembrane domains. This indicates that apLec-1 and apLec-2 probably represent an inhibitory/ activating receptor pair.

The apparent rarity of apLec-2 mRNA is also intriguing. It is only present at detectable levels in the spleen and lung, supporting a role for this molecule in immune surveillance. This is in contrast to the apLec-1 mRNA, which was detectable in all tissues examined. Similar results were seen with DCAR, which was detectable only in the lung, spleen, skin and lymph node, and was less abundantly expressed than the DCIR mRNA (Kanazawa *et al.*, 2003).

4.1.2 Genomic organization of the apLec genes

Attempts to determine the genomic organization of these genes have been unsuccessful thus far. PCR experiments were unable to amplify products from either of these loci, whether attempts were made to amplify the whole gene or merely parts. The difficulty in amplifying these genes imply that they are large genes, which is unsurprising as related mammalian genes are 7-15 kb. It is likely that each of these genes is at least the size of the largest band seen on southern blots. Thus, it is likely that the identical bands seen on these blots are indicative of these genes being on similarly sized fragments as a result of sequence duplication, rather than being adjacent genes on the same restriction fragment. In humans the DCIR and DLEC genes are separated by as much as 463kb (Arce *et al.*, 2001) while in mice DCAR and DCIR are adjacent genes separated by 51.5 kb (Kanazawa *et al.*, 2003), thus it is likely that apLec-1 and apLec-2 are encoded by nearby genes, but that they are not sufficiently close to appear on the same restriction fragment.

The degree of similarity at the cDNA level to the related mammalian genes implies a similar intron/ exon organization and possibly a similar genomic organization. DCIR and DCAR are linked genes located near to, but outside of the NKC in both humans and mice. An avian NKC has yet to be identified and avian genes resembling those found in the NKC of mammals that have been identified thus far have been mapped to the MHC region. In chickens, BNK and B-Lec have been mapped to the *Mhc B* locus and Y-Lec1 and Y-Lec2 have been identified in the Rfp-Y locus. However, these lectins are more closely related to the NKR-P1 family of genes and AICL, CD69 and the CLR family of genes respectively than they are to the lectins described here. Thus, mapping the apLec-1 and apLec-2 genes and identifying flanking loci may uncover an avian NKC.

4.1.3 Splicing of the apLec genes

Alternative splicing is common to many lectin-encoding genes including DC-SIGN, Dectin-1 and DCIR, thus the existence of multiple splice forms for apLec-2 is unsurprising. The transmembrane deletion variant identified by RT-PCR is particularly reminiscent of that observed for DCIR (Huang *et al.*, 2001; Richard *et al.*, 2003). It has been proposed that these variants may be a means of regulating the amount of receptor at the cell surface. In neutrophils the TM-less variant of the DCIR message accumulates in response to certain activation stimuli, which accompanies reduced surface expression of the protein (Richard *et al.*, 2002). The splice variants of apLec-2 may similarly exist to limit the amount of apLec-2 expressed at the cell surface.

The clone 9H3 appears to be an aberrant transcript. The clone features two overlapping ORFs, one of which encodes a CRD similar to those of apLec-2 and apLec-1. The reading frame appears to have been altered by a frameshift, however RT-PCR

experiments failed to identify any other forms of this message, nor have any other copies of this message been identified in the cDNA library. 9H3 may represent an incorrectly spliced message, similar to clone 16F11, but this seems unlikely, as RT-PCR experiments should have identified the correctly spliced message. It is possible that the primers used fall in a position that is altered in the correctly spliced form, though this is unlikely given the placement of the primers relative to the apparent frameshift. Another possibility is that the correctly spliced form is extremely rare though this too is unlikely given the sensitivity of RT-PCR. Another possibility is that 9H3 is expressed from a non-functional pseudogene similar to human Ly49L.

4.1.4 Possible functions of the apLec proteins

The function of apLec-1 and apLec-2 remain unknown, as is the case for DCIR and DCAR, however the significant similarities between them suggest that their functions are homologous. The expression pattern of the apLec messages suggests that they are expressed by cells capable of antigen presentation. Preliminary data indicates that, like mammalian DCIR, expression of the apLec genes is downregulated in response to maturation stimuli, specifically LPS. Taken together these data indicate that these receptors play a role in recognition events important to APCs that are in the antigen uptake phase, such as immature DCs. Thus, it can be imagined that the inhibitory receptor apLec-1, upon binding a ligand such as extracellular self-glycoproteins inhibits activation/ maturation events in response to that ligand. On the other hand, apLec-2, acting as an activating receptor, is likely to induce activation or maturation events in response to a distinct, but structurally related ligand such as self-glycoproteins that are usually intracellular.

Both the primary structures, as well as the predicted tertiary structures support the possibility that apLec-1 and apLec-2 bind distinct ligands. Despite the high level of identity between the CRDs of these receptors, the differences between them suggest that distinct ligands are bound by these receptors. For example, the insertion of two amino acids in the loop between α-helix 2 and β-strand 2 of apLec-1 is a small but significant change that results in a difference in the predicted structure of the putative ligand binding region of these receptors and thus, do likely reflect functional disparities between them. Also, most of the amino acid differences between the two proteins result in changes in charge or hydrophobicity at those sites, which are likely to have an impact on the binding properties of the whole CRD. Also, glycosylation of the lectin itself has been shown to be a determinant of ligand specificity (Marshall and Gordon, 2004). Thus, the more extensive glycosylation of the CRD of apLec-1 compared to that of apLec-2 is also likely to confer differing ligand specificities on these two related receptors.

4.1.5 Significance of identifying potential duck APC lectins

Ducks are important model organisms in the study of infections relevant to man. They are one of very few natural model organisms available for the study of Hepatitis B virus infection and they are the natural reservoir of Influenza A virus. Despite their obvious importance as model organisms, relatively little is known about the immune response, and particularly cell mediated immunity, in these organisms. Without such information, it is difficult to extrapolate experimental data from model infections to what may occur in man. This is especially relevant in vaccine trials, where the success or failure of a given vaccination strategy in ducks may be due to peculiarities of the duck immune response rather than to the strength or weakness of the strategy itself. Currently

there are essentially no markers for distinct leukocyte populations in ducks. Thus, at minimum, the discovery of lectins that are likely to be associated with specific cell types or activation states will provide markers for these cells. For example, apLec-1 and apLec-2 may prove to be effective markers of antigen presenting cells in the antigen uptake phase, and as such, will be useful tools in the study of antigen presentation in ducks. To that end, work is underway to express these proteins and to generate antibodies against them, which should aid in the characterization of duck APCs and provide additional clues as to the function of these proteins. Ideally, these receptors may be useful as more than just markers. Their immunomodulatory potential could be exploited in targeted vaccine therapy. Delivering Hepatitis B vaccines directly to DCs in a manner that would trigger an activating receptor may be a useful way to break tolerance to HBV in chronically infected individuals.

4.2.1 Duck AICL-like genes

The second group of genes that we have identified as members of the C-type lectin superfamily show several features of proteins encoded by genes within the NKC. Like most of those mammalian proteins, these duck proteins have a potential transmembrane region, but lack an N-terminal signal sequence suggesting a type II transmembrane orientation. They have relatively short N-terminal cytoplasmic domains, and a single C-terminal C-type lectin-like domain that deviates from classical C-type lectin CRDs in that they lack the residues involved in calcium ligation. Several features of these duck lectins indicate that they are related to a group of mammalian NKC encoded genes related to, and physically near to, CD69. This group includes CD69 itself, LLT1 and AICL in humans and the CLR family of genes in mice. This group of

mammalian genes and the duck genes are divided into five exons rather than six, as is typical of other NKC encoded lectins. They lack a discrete neck-region encoding exon that is evident in genes like CD94 and the NKR-P1 and LY49 families. As a result, the proteins encoded by these genes have relatively short neck regions: in the case of the two proteins belonging to this group that have been identified thus far, this neck region is only twelve amino acids in length. These duck proteins, like the CD69 related proteins in mammals, lack either an ITIM motif or a charged residue that would allow them to associate with adaptor molecules. Many NKC encoded proteins are expressed exclusively by NK cells or by NK cells and subsets of cytotoxic T cells, however CD69 and the related genes nearby are expressed more widely, in many different cells of hematopoietic lineage. The broad distribution of the duck AICL-like messages certainly suggests that this is true for these genes as well.

The AICL-like genes identified thus far are likely to be expressed from two loci. The four transcripts within the AICL-like family that feature identical ORFs are likely to be expressed from the same locus. 33E5, which has the most divergent UTR, likely represents one allele while 14D12, 22G6 and 29D1 are all transcripts from the other allele, with the differences in their UTRs the results of transcription errors or an artifact of reverse transcription. 29D1 is likely the most complete message from this allele followed by 14D12. The deletion in the 3′ UTR of 22G6 may be a result of a transcriptional error such as template slippage that was included in the cDNA library. Genomic data suggests that clone 6D9 on the other hand represents a distinct locus.

AICL-like 1 and AICL-like 2 are 89.9% identical, but the differences between them are potentially significant. The amino acid differences in the cytoplasmic domain

of these two proteins are not conservative: several positively charged residues in AICL-like 1 are replaced by uncharged residues or in one case a negatively charged residue in AICL-like 2, and most of the remaining substitutions replace polar amino acids with hydrophobic ones or vice versa. These differences are likely to have functional consequences. For example, the second serine in the SSPH motif in the cytoplasmic domain of AICL-like 2 is more likely to be phosphorylated than the serine in the FSPH motif of AICL-like 1 according to the Net Phos prediction. Whether this predicted difference in phosphorylation efficiency is true *in vivo* and of functional importance remains to be determined. The CRDs of these two proteins also showed relatively few differences, but they are substantial ones. Of nine amino acid differences, only a phenylalanine to tyrosine substitution represents a conservative change. Five negatively charged residues in the CRD of AICL-like 1 are absent from the CRD of AICL-like 2 and the remaining changes replace hydrophobic residues from AICL-like 1 with charged or polar residues at equivalent positions in AICL-like 2. As was the case with the cytoplasmic domain, the amino acid differences between the CRDs of these two proteins also result in differences in predicted post-translational modifications. An asparagine residue predicted to be glycosylated in AICL-like 2 is replaced with an aspartate in AICL-like 1, thus only AICL-like 2 is likely to undergo any N-linked glycosylation. Due to this and the charge differences in the CRD of these proteins they are not likely to have the same binding characteristics, such that they may either recognize different ligands or bind to their ligand with different affinities. These possibilities deserve exploration. These proteins should be expressed and purified *in vitro* to study their biochemical

properties and determine their ligands in order to address the impact of the differences between these proteins and determine their biological function.

4.2.2 Expression of duck AICL-like genes

These lectins are expressed in more tissues and more abundantly than would be expected based on their similarity to mammalian activation induced lectins. This broad expression is reminiscent of CLRB (Plougastel *et al.,* 2001). However, the expression pattern seen may be due to the fact that the duck in which their tissue expression was studied was congenitally infected with Hepatitis B virus or that individual may have been infected with another pathogen leading to a broad expression of the AICL-like genes. This expression pattern should be confirmed in an uninfected duck to determine whether HBV infection had any effect on the expression of these lectins. Analyzing the promoters of these genes should also clarify the mechanism by which they are regulated and the cells and tissues in which they should be expressed under normal circumstances.

4.2.3 Potential Functions of the AICL-like receptors

These AICL-like duck receptors are related to the activation antigens CD69 and AICL, and thus may be activation antigens themselves. The predicted AICL-like proteins share the highest sequence identity with the CRD of human AICL, yet bear structural features, such, as a larger cytoplasmic domain, that are more similar to CD69. Thus, these genes may more closely resemble the ancestral gene that was duplicated and diverged to form the AICL and CD69 genes in man.

Both CD69 and AICL are rapidly induced upon cellular activation events. The significance of AICL induction and its function remain unknown. CD69, on the other hand, is capable of transmitting signals involved in the induction of certain genes, such as

the IL-2 receptor α gene, and of triggering effector functions of the cells that express it, for example cytotoxicity in NK cells.

Several features shared between CD69 and the duck AICL-like lectins, but not AICL, imply that they may be capable of fulfilling a similar role to that of CD69. First, there is the longer cytoplasmic domain. The cytoplasmic domain of CD69 is of a similar length to the duck AICL-like proteins and is constitutively phosphorylated. The presence of serines that are predicted to be phosphorylated, especially the serine within the SPH motif conserved between the cytoplasmic domains of the duck AICL-like lectins and human CD69, suggest functional similarity to CD69. Second, the predicted tertiary structure of the duck AICL-like proteins and that known for the CRD of CD69 are also highly similar despite only moderately identical primary structures. Third, of the properties these receptors share with CD69, the potential, through post-translational modification, to be recruited to lipid rafts is perhaps the most significant. Lipid rafts, or detergent resistant microdomains, are important "platforms" for various signaling events. It has been proposed that CD69 may mediate signaling by association with adaptor molecules through co-localization to lipid rafts rather than by the charge complementarity mechanism employed by related receptors such as the activating members of the Ly49 family. The duck AICL proteins could also be recruited to lipid rafts. As many as three cysteines within the cytoplasmic domain and the N-terminal portion of the transmembrane domain may be palmitoylated in both of the AICL-like proteins. Also, despite deviating from the consensus sequence (M G X X X S/T Y/R/X Y/R/X), these proteins may also be myristoylated, as they do feature the absolutely required glycine at position 2 and many proteins with even more pronounced deviations from the consensus

motif are known to be myristoylated (Resh, 1999). These modifications should act to recruit the duck proteins into lipid rafts were they could then associate with adaptor proteins such as LAT (Zhang *et al.*, 1998). In fact the relative positions of potential acylation and phosphorylation sites between CD69, the duck AICL proteins and LAT is similar, implying a similar conformation of these potentially functionally significant features of their cytoplasmic domains. The topology may be important in mediating association between the receptor and adaptor. Thus, despite the lack of intrinsic signaling domains, these proteins may represent signaling competent receptors.

These lectins also show relatively high homology to the human LLT1 and the murine CLR family of genes. However, unlike the CLR and LLT1 encoded proteins, these lectins retain all six cysteines typical of a CRD, whereas CLR family members are missing either C5 or both C4 and C5 of the CRD (Plougastel *et al.,* 2001). Despite this potentially significant difference, without some functional data, we cannot discount the possibility that these duck lectins are in fact CLR homologs.

It has recently been demonstrated that CLRB and CLRG are ligands for NKR-P1d and NKR-P1f respectively, which are encoded by nearby genes (Iizuka *et al.*, 2003). Thus, one possibility is that these duck genes may encode ligands for duck NKR-P1 homologs. In chickens, the BNK gene, which is related most closely to the NKRP1 genes, is adjacent to the CLR-like B-Lec (Kaufman *et al.*, 1999), which is also quite similar to the duck genes described here, and may represent a similar case of a genetically linked receptor-ligand pair. However, attempts to uncover a BNK/ NKRP1 homolog in the duck have been unsuccessful thus far.

The high degree of similarity between these duck proteins and chicken MHC-linked lectin-like immunoreceptors raises the possibility that these genes may be localized to the MHC region in ducks rather than being found in a region of conserved synteny to the mammalian NKC. The absence of lectins in the MHC region of teleost fish implies that this translocation of lectin-like genes to the MHC may be unique to the avian lineage and that the maintenance of this condition may provide a selective advantage. As these receptors are not closely related to the MHC recognizing receptors such as CD94, NKG2 or Ly49 it seems unlikely that this linkage allows for co-evolution of these receptors with their ligand, which is the most obvious advantage such linkage could provide. However, the recent discovery that the intertwined NKR-P1 and CLR families of lectins in fact represent receptor-ligand pairs may be reflected in the linked homologous genes in the form of B-Lec and B-NK within the chicken MHC region. Thus the presence of these genes in the *Mhc* may be seen as the co-localization of genes encoding markers of self that inhibit NK cell killing, and a receptor that detects them, to one region. Thus, the loci represented by 6D9 and 14D12 are likely homologous MHC-linked lectins. In light of this hypothesis, it is intriguing that we have yet to identify a duck B-NK homolog. While these lectins resemble those found in the MHC of other avian species the restriction patterns observed for these duck lectins is not consistent with that of the classical MHC class I genes, in as much as not all ducks with the same restriction pattern for the AICL-like family of genes have the same restriction fragment pattern for MHC class I. This may mean that these genes are in an NKC-like locus or that they are linked to an Rfp-Y-like locus that segregates independently of the MHC class I genes.

In chickens resistance or susceptibility to Marek's Disease Virus has been linked to the *Mhc B* region, however this variation in levels of resistance in some B locus haplotypes is not due to differences in MHC class I genes and may be connected to either the B-NK or B-Lec alleles associated with a particular haplotype (Kaufman, 2000). Given the similarity between the duck AICL genes we have cloned and the chicken B-Lec and Y-Lec genes, they are likely to have similar importance in determining the antiviral immune responses of ducks. Given the finding that these are polymorphic genes, determining the contribution of these lectins to the immune response may illuminate a key factor contributing to inter-individual variations in immune responsiveness to viral challenge in ducks.

5 Bibliography:

AHN, J. S., HAMILTON, M. J., DAVIS, W. C., and PARK, Y. H. (2002). Cloning of bovine CD69. *Vet Immunol & Immunopathol* **88**(1-2), 43-8.

ALVAREZ, C. P., LASALA, F., CARRILLO, J., MUNIZ, O., CORBI, A. L., and DELGADO, R. (2002). C-type lectins DC-SIGN and L-SIGN mediate cellular entry by Ebola virus in cis and in trans. *Journal of Virology* **76**(13), 6841-4.

AMAKATA, Y., FUJIYAMA, Y., ANDOH, A., HODOHARA, K., and BAMBA, T. (2001). Mechanism of NK cell activation induced by coculture with dendritic cells derived from peripheral blood monocytes. *Clin Exp Immunol* **124**(2), 214-22.

ARAGANE, Y., MAEDA, A., SCHWARZ, A., TEZUKA, T., ARIIZUMI, K., and SCHWARZ, T. (2003). Involvement of dectin-2 in ultraviolet radiation-induced tolerance. *J Immuno.* **171**(7), 3801-7.

ARASE, H., MOCARSKI, E. S., CAMPBELL, A. E., HILL, A. B., and LANIER, L. L. (2002). Direct recognition of cytomegalovirus by activating and inhibitory NK cell receptors. *Science* **296**(5571), 1323-6.

ARCE, I., MARTINEZ-MUNOZ, L., RODA-NAVARRO, P., and FERNANDEZ-RUIZ, E. (2004). The human C-type lectin CLECSF8 is a novel monocyte/macrophage endocytic receptor. Eur *J Immunol* **34**(1), 210-220.

ARCE, I., RODA-NAVARRO, P., MONTOYA, M. C., HERNANZ-FALCON, P., PUIG-KROGER, A., and FERNANDEZ-RUIZ, E. (2001). Molecular and genomic characterization of human DLEC, a novel member of the C-type lectin receptor gene family preferentially expressed on monocyte-derived dendritic cells. *Eur J Immunol* **31**(9), 2733-40.

ARIIZUMI, K., SHEN, G. L., SHIKANO, S., RITTER, R., 3RD, ZUKAS, P., EDELBAUM, D., MORITA, A., and TAKASHIMA, A. (2000a). Cloning of a second dendritic cell-associated C-type lectin (dectin-2) and its alternatively spliced isoforms. *J Biol Chem* **275**(16), 11957-63.

ARIIZUMI, K., SHEN, G. L., SHIKANO, S., XU, S., RITTER, R., 3RD, KUMAMOTO, T., EDELBAUM, D., MORITA, A., BERGSTRESSER, P. R., and TAKASHIMA, A. (2000b). Identification of a novel, dendritic cell-associated molecule, dectin- 1, by subtractive cDNA cloning. *J Biol Chem* **275**(26), 20157-67.

BALCH, S. G., GREAVES, D. R., GORDON, S., and MCKNIGHT, A. J. (2002). Organization of the mouse macrophage C-type lectin (Mcl) gene and identification of a subgroup of related lectin molecules. *Eur J Immunogenet* **29**(1), 61-4.

BALCH, S. G., MCKNIGHT, A. J., SELDIN, M. F., and GORDON, S. (1998). Cloning of a novel C-type lectin expressed by murine macrophages. *J Biol Chem.* **273**(29), 18656-64.

BASHIROVA, A. A., GEIJTENBEEK, T. B., VAN DUIJNHOVEN, G. C., VAN VLIET, S. J., EILERING, J. B., MARTIN, M. P., WU, L., MARTIN, T. D., VIEBIG, N., KNOLLE, P. A., KEWALRAMANI, V. N., VAN KOOYK, Y., and CARRINGTON, M. (2001). A dendritic cell-specific intercellular adhesion molecule 3-grabbing nonintegrin (DC-SIGN)-related protein is highly expressed on human liver sinusoidal endothelial cells and promotes HIV-1 infection. *J Exp Med* **193**(6), 671-8.

BATES, E. E., FOURNIER, N., GARCIA, E., VALLADEAU, J., DURAND, I., PIN, J. J., ZURAWSKI, S. M., PATEL, S., ABRAMS, J. S., LEBECQUE, S., GARRONE, P., and SAELAND, S. (1999). APCs express DCIR, a novel C-type lectin surface receptor containing an immunoreceptor tyrosine-based inhibitory motif. *J Immunol* **163**(4), 1973-83.

BERNOT, A., ZOOROB, R., and AUFFRAY, C. (1994). Linkage of a new member of the lectin supergene family to chicken Mhc genes. *Immunogenetics* **39**(4), 221-9.

BEZOUSKA, K., NEPOVIM, A., HORVATH, O., POSPISIL, M., HAMANN, J., and FEIZI, T. (1995). CD 69 antigen of human lymphocytes is a calcium-dependent carbohydrate-binding protein. *Biochem & Biophys Res Comm* **208**(1), 68-74.

BLOM, N., GAMMELTOFT, S., and BRUNAK, S. (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* **294**(5), 1351-62.

BOLES, K. S., BARTEN, R., KUMARESAN, P. R., TROWSDALE, J., and MATHEW, P. A. (1999). Cloning of a new lectin-like receptor expressed on human NK cells. *Immunogenetics* **50**(1-2), 1-7.

BONIFAZ, L., BONNYAY, D., MAHNKE, K., RIVERA, M., NUSSENZWEIG, M. C., and STEINMAN, R. M. (2002). Efficient targeting of protein antigen to the dendritic cell receptor DEC-205 in the steady state leads to antigen presentation on major histocompatibility complex class I products and peripheral CD8+ T cell tolerance. *J Exp Med* **196**(12), 1627-38.

BORREGO, F., ROBERTSON, M. J., RITZ, J., PENA, J., and SOLANA, R. (1999). CD69 is a stimulatory receptor for natural killer cell and its cytotoxic effect is blocked by CD94 inhibitory receptor. *Immunology* **97**(1), 159-65.

BROWN, G. D., and GORDON, S. (2001). Immune recognition. A new receptor for beta-glucans. *Nature* **413**(6851), 36-7.

BROWN, G. D., HERRE, J., WILLIAMS, D. L., WILLMENT, J. A., MARSHALL, A. S., and GORDON, S. (2003). Dectin-1 mediates the biological effects of beta-glucans. *J Exp Med* **197**(9), 1119-24.

BROWN, G. D., TAYLOR, P. R., REID, D. M., WILLMENT, J. A., WILLIAMS, D. L., MARTINEZ-POMARES, L., WONG, S. Y., and GORDON, S. (2002). Dectin-1 is a major beta-glucan receptor on macrophages. *J Exp Med* **196**(3), 407-12.

BROWN, M. G., SCALZO, A. A., MATSUMOTO, K., and YOKOYAMA, W. M. (1997). The natural killer gene complex: a genetic basis for understanding natural killer cell function and innate immunity. *Immunol Rev* **155**, 53-65.

BYRNE, S. N., and HALLIDAY, G. M. (2002). Dendritic cells: making progress with tumour regression? *Immunol Cell Biol* **80**(6), 520-30.

CARLYLE, J. R., MARTIN, A., MEHRA, A., ATTISANO, L., TSUI, F. W., and ZUNIGA-PFLUCKER, J. C. (1999). Mouse NKR-P1B, a novel NK1.1 antigen with inhibitory function. *J Immunol* **162**(10), 5917-23.

CAUX, C., MASSACRIER, C., VANBERVLIET, B., DUBOIS, B., VAN KOOTEN, C., DURAND, I., and BANCHEREAU, J. (1994). Activation of human dendritic cells through CD40 cross-linking. *J Exp Med* **180**(4), 1263-72.

CEBRIAN, M., YAGUE, E., RINCON, M., LOPEZ-BOTET, M., DE LANDAZURI, M. O., and SANCHEZ-MADRID, F. (1988). Triggering of T cell proliferation through AIM, an activation inducer molecule expressed on activated human lymphocytes. *J Exp Med* **168**(5), 1621-37.

CHEHIMI, J., LUO, Q., AZZONI, L., SHAWVER, L., NGOUBILLY, N., JUNE, R., JERANDI, G., FARABAUGH, M., and MONTANER, L. J. (2003). HIV-1 transmission and cytokine-induced expression of DC-SIGN in human monocyte-derived macrophages. *J Leukoc Biol* **74**(5), 757-63.

COLONNA, M., SAMARIDIS, J., and ANGMAN, L. (2000). Molecular characterization of two novel C-type lectin-like receptors, one of which is selectively expressed in human dendritic cells. *Eur J Immunol* **30**(2), 697-704.

COMBET, C., JAMBON, M., DELEAGE, G., and GEOURJON, C. (2002). Geno3D: automatic comparative molecular modelling of protein. *Bioinformatics* **18**(1), 213-4.

DE MARIA, R., CIFONE, M. G., TROTTA, R., RIPPO, M. R., FESTUCCIA, C., SANTONI, A., and TESTI, R. (1994). Triggering of human monocyte activation through CD69, a member of the natural killer cell gene complex family of signal transducing receptors. *J Exp Med* **180**(5), 1999-2004.

DE SMEDT, T., PAJAK, B., MURAILLE, E., LESPAGNARD, L., HEINEN, E., DE BAETSELIER, P., URBAIN, J., LEO, O., and MOSER, M. (1996). Regulation of dendritic cell numbers and maturation by lipopolysaccharide in vivo. *J Exp Med.* **184**(4), 1413-24.

DENDA-NAGAI, K., KUBOTA, N., TSUIJI, M., KAMATA, M., and IRIMURA, T. (2002). Macrophage C-type lectin on bone marrow-derived immature dendritic cells is involved in the internalization of glycosylated antigens. *Glycobiology* **12**(7), 443-50.

DRICKAMER, K. (1988). Two distinct classes of carbohydrate-recognition domains in animal lectins. *J Biol Chem* **263**(20), 9557-60.

DRICKAMER, K. (1992). Engineering galactose-binding activity into a C-type mannose-binding protein. *Nature* **360**(6400), 183-6.

DRICKAMER, K. (1993). CA2+-Dependent carbohydrate-recognition domains in animal proteins. *Curr Opin Struct Biol* **3**(3), 393-400.

DRICKAMER, K., and TAYLOR, M. E. (1993). Biology of animal lectins. *Annu Rev Cell Bio* **9**, 237-64.

DZIONEK, A., SOHMA, Y., NAGAFUNE, J., CELLA, M., COLONNA, M., FACCHETTI, F., GUNTHER, G., JOHNSTON, I., LANZAVECCHIA, A., NAGASAKA, T., OKADA, T., VERMI, W., WINKELS, G., YAMAMOTO, T., ZYSK, M., YAMAGUCHI, Y., and SCHMITZ, J. (2001). BDCA-2, a novel plasmacytoid dendritic cell-specific type II C-type lectin, mediates antigen capture and is a potent inhibitor of interferon alpha/beta induction. *J Exp Med* **194**(12), 1823-34.

EICHLER, W., RUSCHPLER, P., WOBUS, M., and DROSSLER, K. (2001). Differentially induced expression of C-type lectins in activated lymphocytes. *J Cell Biochem* **81**(S36), 201-208.

ENGERING, A., GEIJTENBEEK, T. B., VAN VLIET, S. J., WIJERS, M., VAN LIEMPT, E., DEMAUREX, N., LANZAVECCHIA, A., FRANSEN, J., FIGDOR, C. G., PIGUET, V., and VAN KOOYK, Y. (2002a). The dendritic cell-specific adhesion receptor DC-SIGN internalizes antigen for presentation to T cells. *J Immunol* **168**(5), 2118-26.

ENGERING, A., VAN VLIET, S. J., GEIJTENBEEK, T. B., and VAN KOOYK, Y. (2002b). Subset of DC-SIGN(+) dendritic cells in human blood transmits HIV-1 to T lymphocytes. *Blood* **100**(5), 1780-6.

ESCRIBANO, L., ORFAO, A., VILLARRUBIA, J., MARTIN, F., MADRUGA, J. I., CUEVAS, M., VELASCO, J. L., RIOS, A., and SAN MIGUEL, J. F. (1997). Sequential immunophenotypic analysis of mast cells in a case of systemic mast cell disease evolving to a mast cell leukemia. *Cytometry* **30**(2), 98-102.

FEINBERG, H., MITCHELL, D. A., DRICKAMER, K., and WEIS, W. I. (2001). Structural basis for selective recognition of oligosaccharides by DC-SIGN and DC-SIGNR. *Science* **294**(5549), 2163-6.

FERNANDES, M. J., FINNEGAN, A. A., SIRACUSA, L. D., BRENNER, C., ISCOVE, N. N., and CALABRETTA, B. (1999). Characterization of a novel receptor that maps near the natural killer gene complex: demonstration of carbohydrate binding and expression in hematopoietic cells. *Cancer Res.* **59**(11), 2709-17.

FOERSTER, M., HAEFNER, D., and KROEGEL, C. (2002). Bcl-2-mediated regulation of CD69-induced apoptosis of human eosinophils: identification and characterization of a novel receptor-induced mechanism and relationship to CD95-transduced signalling. *Scand J Immunol* **56**(4), 417-28.

FRISON, N., TAYLOR, M. E., SOILLEUX, E., BOUSSER, M. T., MAYER, R., MONSIGNY, M., DRICKAMER, K., and ROCHE, A. C. (2003). Oligolysine-based oligosaccharide clusters: selective recognition and endocytosis by the mannose receptor and dendritic cell-specific intercellular adhesion molecule 3 (ICAM-3)-grabbing nonintegrin. *J Biol Chem* **278**(26), 23922-9.

GANTNER, B. N., SIMMONS, R. M., CANAVERA, S. J., AKIRA, S., and UNDERHILL, D. M. (2003). Collaborative induction of inflammatory responses by dectin-1 and Toll-like receptor 2. *J Exp Med* **197**(9), 1107-17.

GARDNER, J. P., DURSO, R. J., ARRIGALE, R. R., DONOVAN, G. P., MADDON, P. J., DRAGIC, T., and OLSON, W. C. (2003). L-SIGN (CD 209L) is a liver-specific capture receptor for hepatitis C virus. *Proc Natl Acad Sci* **100**(8), 4498-503.

GEIJTENBEEK, T. B., GROOT, P. C., NOLTE, M. A., VAN VLIET, S. J., GANGARAM-PANDAY, S. T., VAN DUIJNHOVEN, G. C., KRAAL, G., VAN OOSTERHOUT, A. J., and VAN KOOYK, Y. (2002). Marginal zone macrophages express a murine homologue of DC-SIGN that captures blood-borne antigens in vivo. *Blood* **100**(8), 2908-16.

GEIJTENBEEK, T. B., KROOSHOOP, D. J., BLEIJS, D. A., VAN VLIET, S. J., VAN DUIJNHOVEN, G. C., GRABOVSKY, V., ALON, R., FIGDOR, C. G., and VAN KOOYK, Y. (2000a). DC-SIGN-ICAM-2 interaction mediates dendritic cell trafficking. *Nat Immunol* **1**(4), 353-7.

GEIJTENBEEK, T. B., KWON, D. S., TORENSMA, R., VAN VLIET, S. J., VAN DUIJNHOVEN, G. C., MIDDEL, J., CORNELISSEN, I. L., NOTTET, H. S., KEWALRAMANI, V. N., LITTMAN, D. R., FIGDOR, C. G., and VAN KOOYK, Y. (2000b). DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell* **100**(5), 587-97.

GEIJTENBEEK, T. B., TORENSMA, R., VAN VLIET, S. J., VAN DUIJNHOVEN, G. C., ADEMA, G. J., VAN KOOYK, Y., and FIGDOR, C. G. (2000c). Identification of DC-SIGN, a novel dendritic cell-specific ICAM-3 receptor that supports primary immune responses. *Cell* **100**(5), 575-85.

GEROSA, F., BALDANI-GUERRA, B., NISII, C., MARCHESINI, V., CARRA, G., and TRINCHIERI, G. (2002). Reciprocal activating interaction between natural killer cells and dendritic cells. *J Exp Med* **195**(3), 327-33.

GOBEL, T. W. (2000). Isolation and analysis of natural killer cells in chickens. *Methods Mol Biol* **121**, 337-45.

GOBEL, T. W., KASPERS, B., and STANGASSINGER, M. (2001). NK and T cells constitute two major, functionally distinct intestinal epithelial lymphocyte subsets in the chicken. *Int Immunol* **13**(6), 757-62.

GUEX, N., and PEITSCH, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**(15), 2714-23.

GUTA, R., JUNG, E., and BRUNAK, S. (2002). Prediction of N-glycosylation sites in human proteins. *In preparation (according to http://www.cbs.dtu.dk/services/NetNGlyc/.)*

HALARY, F., AMARA, A., LORTAT-JACOB, H., MESSERLE, M., DELAUNAY, T., HOULES, C., FIESCHI, F., ARENZANA-SEISDEDOS, F., MOREAU, J. F., and DECHANET-MERVILLE, J. (2002). Human cytomegalovirus binding to DC-SIGN is required for dendritic cell infection and target cell trans-infection. *Immunity* **17**(5), 653-64.

HAMANN, J., FIEBIG, H., and STRAUSS, M. (1993). Expression cloning of the early activation antigen CD69, a type II integral membrane protein with a C-type lectin domain. *J Immunol* **150**(11), 4920-7.

HAMANN, J., MONTGOMERY, K. T., LAU, S., KUCHERLAPATI, R., and VAN LIER, R. A. (1997). AICL: a new activation-induced antigen encoded by the human NK gene complex. *Immunogenetics* **45**(5), 295-300.

HARA, T., JUNG, L. K., BJORNDAHL, J. M., and FU, S. M. (1986). Human T cell activation. III. Rapid induction of a phosphorylated 28 kD/32 kD disulfide-linked early activation antigen (EA 1) by 12-o-tetradecanoyl phorbol-13-acetate, mitogens, and antigens. *J Exp Med* **164**(6), 1988-2005.

HERMANZ-FALCON, P., ARCE, I., RODA-NAVARRO, P., and FERNANDEZ-RUIZ, E. (2001). Cloning of human DECTIN-1, a novel C-type lectin-like receptor gene expressed on dendritic cells. *Immunogenetics* **53**(4), 288-95.

HUANG, X., YUAN, Z., CHEN, G., ZHANG, M., ZHANG, W., YU, Y., and CAO, X. (2001). Cloning and characterization of a novel ITIM containing lectin-like immunoreceptor LLIR and its two transmembrane region deletion variants. *Biochem Biophys Res Commun* **281**(1), 131-40.

IIZUKA, K., NAIDENKO, O. V., PLOUGASTEL, B. F., FREMONT, D. H., and YOKOYAMA, W. M. (2003). Genetically linked C-type lectin-related ligands for the NKRP1 family of natural killer cell receptors. *Nature Immunol* **4**(8), 801-7.

JIANG, W., SWIGGARD, W. J., HEUFLER, C., PENG, M., MIRZA, A., STEINMAN, R. M., and NUSSENZWEIG, M. C. (1995). The receptor DEC-205 expressed by dendritic cells and thymic epithelial cells is involved in antigen processing. *Nature* **375**(6527), 151-5.

KANAZAWA, N., OKAZAKI, T., NISHIMURA, H., TASHIRO, K., INABA, K., and MIYACHI, Y. (2002). DCIR acts as an inhibitory receptor depending on its immunoreceptor tyrosine-based inhibitory motif. *J Invest Dermatol* **118**(2), 261-6.

KANAZAWA, N., TASHIRO, K., INABA, K., and MIYACHI, Y. (2003). Dendritic Cell Immunoactivating Receptor, a Novel C-type Lectin Immunoreceptor, Acts as an Activating Receptor through Association with Fc Receptor {gamma} Chain. *J Biol Chem* **278**(35), 32645-52.

KAUFMAN, J., MILNE, S., GOBEL, T. W., WALKER, B. A., JACOB, J. P., AUFFRAY, C., ZOOROB, R., and BECK, S. (1999). The chicken B locus is a minimal essential major histocompatibility complex. *Nature* **401**(6756), 923-5.

KAUFMAN, J. (2000). The simple chicken major histocompatibility complex: life and death in the face of pathogens and vaccines. Philos Trans R Soc Lond B Biol Sci. **355** (1400), 1077-84.

KOZAK, M. (1997). Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6. *EMBO J* **16**(9), 2482-92.

KUNG, S. K., SU, R. C., SHANNON, J., and MILLER, R. G. (1999). The NKR-P1B gene product is an inhibitory receptor on SJL/J NK cells. *J Immunol* **162**(10), 5876-87.

LANE, P. J., and BROCKER, T. (1999). Developmental regulation of dendritic cell function. *Curr Opin Immunol* **11**(3), 308-13.

LANIER, L. L., BUCK, D. W., RHODES, L., DING, A., EVANS, E., BARNEY, C., and PHILLIPS, J. H. (1988). Interleukin 2 activation of natural killer cells rapidly induces the expression and phosphorylation of the Leu-23 activation antigen. *J Exp Med* **167**(5), 1572-85.

LANIER, L. L., CHANG, C., and PHILLIPS, J. H. (1994). Human NKR-P1A. A disulfide-linked homodimer of the C-type lectin superfamily expressed by a subset of NK and T lymphocytes. *J Immunol* **153**(6), 2417-28.

LI, Y., WANG, M. N., LI, H., KING, K. D., BASSI, R., SUN, H., SANTIAGO, A., HOOPER, A. T., BOHLEN, P., and HICKLIN, D. J. (2002). Active immunization against the vascular endothelial growth factor receptor flk1 inhibits tumor angiogenesis and metastasis. *J Exp Med* **195**(12), 1575-84.

LUFT, T., PANG, K. C., THOMAS, E., HERTZOG, P., HART, D. N., TRAPANI, J., and CEBON, J. (1998). Type I IFNs enhance the terminal differentiation of dendritic cells. *J Immunol.* **161**(4), 1947-53.

MAHNKE, K., GUO, M., LEE, S., SEPULVEDA, H., SWAIN, S. L., NUSSENZWEIG, M., and STEINMAN, R. M. (2000). The dendritic cell receptor for endocytosis, DEC-205, can recycle and enhance antigen presentation via major histocompatibility complex class II-positive lysosomal compartments. *J Cell Biol* **151**(3), 673-84.

MARSHALL, A. S. J., and GORDON, S. (2004). Commentary: C-type lectins on the macrophage cell surface - recent findings. *Eur J Immunol* **34**(1), 18-24.

MATSUMOTO, M., TANAKA, T., KAISHO, T., SANJO, H., COPELAND, N. G., GILBERT, D. J., JENKINS, N. A., and AKIRA, S. (1999). A novel LPS-inducible C-type lectin is a transcriptional target of NF-IL6 in macrophages. *J Immunol* **163**(9), 5039-48.

MITCHELL, D. A., FADDEN, A. J., and DRICKAMER, K. (2001). A novel mechanism of carbohydrate recognition by the C-type lectins DC-SIGN and DC-SIGNR. Subunit organization and binding to multivalent ligands. *J Biol Chem* **276**(31), 28939-45.

MORETTA, A., POGGI, A., PENDE, D., TRIPODI, G., ORENGO, A. M., PELLA, N., AUGUGLIARO, R., BOTTINO, C., CICCONE, E., and MORETTA, L. (1991). CD69-mediated pathway of lymphocyte activation: anti-CD69 monoclonal antibodies trigger the cytolytic activity of different lymphoid effector cells with the exception of cytolytic T lymphocytes expressing T cell receptor alpha/beta. *J Exp Med* **174**(6), 1393-8.

MUMMIDI, S., CATANO, G., LAM, L., HOEFLE, A., TELLES, V., BEGUM, K., JIMENEZ, F., AHUJA, S. S., and AHUJA, S. K. (2001). Extensive repertoire of membrane-bound and soluble dendritic cell-specific ICAM-3-grabbing nonintegrin 1 (DC-SIGN1) and DC-SIGN2 isoforms. Inter-individual variation in expression of DC-SIGN transcripts. *J Biol Chem* **276**(35), 33196-212.

NAKAMURA, S., SUNG, S. S., BJORNDAHL, J. M., and FU, S. M. (1989). Human T cell activation. IV. T cell activation and proliferation via the early activation antigen EA 1. *J Exp Med* **169**(3), 677-89.
NATARAJAN, K., SAWICKI, M. W., MARGULIES, D. H., and MARIUZZA, R. A. (2000). Crystal structure of human CD69: a C-type lectin-like activation marker of hematopoietic cells. *Biochemistry* **39**(48), 14779-86.

NGUYEN, D. G., and HILDRETH, J. E. (2003). Involvement of macrophage mannose receptor in the binding and transmission of HIV by macrophages. *EurJ Immunol* **33**(2), 483-93.

PARENT, S. A., ZHANG, T., CHREBET, G., CLEMAS, J. A., FIGUEROA, D. J., KY, B., BLEVINS, R. A., AUSTIN, C. P., and ROSEN, H. (2002). Molecular characterization of the murine SIGNR1 gene encoding a C-type lectin homologous to human DC-SIGN and DC-SIGNR. *Gene* **293**(1-2), 33-46.

PARK, C. G., TAKAHARA, K., UMEMOTO, E., YASHIMA, Y., MATSUBARA, K., MATSUDA, Y., CLAUSEN, B. E., INABA, K., and STEINMAN, R. M. (2001). Five mouse homologues of the human dendritic cell C-type lectin, DC-SIGN. *Internat Immunol* **13**(10), 1283-90.

PAVLICEK, J., SOPKO, B., ETTRICH, R., KOPECKY, V., JR., BAUMRUK, V., MAN, P., HAVLICEK, V., VRBACKY, M., MARTINKOVA, L., KREN, V., POSPISIL, M., and BEZOUSKA, K. (2003). Molecular characterization of binding of calcium and carbohydrates by an early activation antigen of lymphocytes CD69. *Biochemistry* **42**(31), 9295-306.

PICCIOLI, D., SBRANA, S., MELANDRI, E., and VALIANTE, N. M. (2002). Contact-dependent stimulation and inhibition of dendritic cells by natural killer cells. *J Exp Med* **195**(3), 335-41.

PISEGNA, S., ZINGONI, A., PIROZZI, G., CINQUE, B., CIFONE, M. G., MORRONE, S., PICCOLI, M., FRATI, L., PALMIERI, G., and SANTONI, A. (2002). Src-Dependent Syk Activation Controls CD69-Mediated Signaling and Function on Human NK Cells. *J Immunol* **169**(1), 68-74.

PLOUGASTEL, B., DUBBELDE, C., and YOKOYAMA, W. M. (2001). Cloning of Clr, a new family of lectin-like genes localized between mouse Nkrp1a and Cd69. *Immunogenetics* **53**(3), 209-14.

PLOUGASTEL, B., MATSUMOTO, K., DUBBELDE, C., and YOKOYAMA, W. M. (2001). Analysis of a 1-Mb BAC contig overlapping the mouse Nkrp1 cluster of genes: cloning of three new Nkrp1 members, Nkrp1d, Nkrp1e, and Nkrp1f. *Immunogenetics* **53**(7), 592-8.

POGGI, A., RUBARTELLI, A., MORETTA, L., and ZOCCHI, M. R. (1997). Expression and function of NKRP1A molecule on human monocytes and dendritic cells. *Eur J Immunol* **27**(11), 2965-70.

POGGI, A., ZOCCHI, M. R., COSTA, P., FERRERO, E., BORSELLINO, G., PLACIDO, R., GALGANI, S., SALVETTI, M., GASPERINI, C., RISTORI, G., BROSNAN, C. F., and BATTISTINI, L. (1999). IL-12-mediated NKRP1A up-regulation and consequent enhancement of endothelial transmigration of V delta 2+ TCR gamma delta+ T lymphocytes from healthy donors and multiple sclerosis patients. *J Immunol* **162**(7), 4349-54.

POHLMANN, S., LESLIE, G. J., EDWARDS, T. G., MACFARLAN, T., REEVES, J. D., HIEBENTHAL-MILLOW, K., KIRCHHOFF, F., BARIBAUD, F., and DOMS, R. W. (2001). DC-SIGN interactions with human immunodeficiency virus: virus binding and transfer are dissociable functions. *J Virol* **75**(21), 10523-6.

POHLMANN, S., ZHANG, J., BARIBAUD, F., CHEN, Z., LESLIE, G. J., LIN, G., GRANELLI-PIPERNO, A., DOMS, R. W., RICE, C. M., and MCKEATING, J. A. (2003). Hepatitis C virus glycoproteins interact with DC-SIGN and DC-SIGNR. *J Virol* **77**(7), 4070-80.

RENEDO, M., ARCE, I., RODRIGUEZ, A., CARRETERO, M., LANIER, L. L., LOPEZ-BOTET, M., and FERNANDEZ-RUIZ, E. (1997). The human natural killer gene complex is located on chromosome 12p12- p13. *Immunogenetics* **46**(4), 307-11.
RESH, M. D. (1999). Fatty acylation of proteins: new insights into membrane targeting of myristoylated and palmitoylated proteins. *Biochim Biophys Acta* **1451**(1), 1-16.

RICHARD, M., THIBAULT, N., VEILLEUX, P., BRETON, R., and BEAULIEU, A. D. (2003). The ITIM-bearing CLECSF6 (DCIR) is down-modulated in neutrophils by neutrophil activating agents. *Biochem & Biophys Res Comm* **310**(3), 767-73.

RICHARD, M., VEILLEUX, P., ROULEAU, M., PAQUIN, R., and BEAULIEU, A. D. (2002). The expression pattern of the ITIM-bearing lectin CLECSF6 in neutrophils suggests a key role in the control of inflammation. *J Leukoc Biol.* **71**(5), 871-880.

RISSO, A., SMILOVICH, D., CAPRA, M. C., BALDISSARRO, I., YAN, G., BARGELLESI, A., and COSULICH, M. E. (1991). CD69 in resting and activated T lymphocytes. Its association with a GTP binding protein and biochemical requirements for its expression. *J Immunol* **146**(12), 4105-14.

ROGERS, S., SHAW, I., ROSS, N., NAIR, V., ROTHWELL, L., KAUFMAN, J., and KAISER, P. (2003). Analysis of part of the chicken Rfp-Y region reveals two novel lectin genes, the first complete genomic sequence of a class I alpha-chain gene, a

truncated class II beta-chain gene, and a large CR1 repeat. *Immunogenetics* **12,** 12.

RYAN, E. J., MARSHALL, A. J., MAGALETTI, D., FLOYD, H., DRAVES, K. E., OLSON, N. E., and CLARK, E. A. (2002). Dendritic cell-associated lectin-1: a novel dendritic cell-associated, C-type lectin-like molecule enhances T cell secretion of IL-4. *J Immunol* **169**(10), 5638-48.

SAIKH, K. U., KHAN, A. S., KISSNER, T., and ULRICH, R. G. (2001). IL-15-induced conversion of monocytes to mature dendritic cells. *Clin Exp Immunol.* **126**(3), 447-55.

SALLUSTO, F., CELLA, M., DANIELI, C., and LANZAVECCHIA, A. (1995). Dendritic cells use macropinocytosis and the mannose receptor to concentrate macromolecules in the major histocompatibility complex class II compartment: downregulation by cytokines and bacterial products. *J Exp Med* **182**(2), 389-400.

SANCHO, D., SANTIS, A. G., ALONSO-LEBRERO, J. L., VIEDMA, F., TEJEDOR, R., and SANCHEZ-MADRID, F. (2000). Functional analysis of ligand-binding and signal transduction domains of CD69 and CD23 C-type lectin leukocyte receptors. *J Immunol* **165**(7), 3868-75.

SANTIS, A. G., LOPEZ-CABRERA, M., HAMANN, J., STRAUSS, M., and SANCHEZ-MADRID, F. (1994). Structure of the gene coding for the human early lymphocyte activation antigen CD69: a C-type lectin receptor evolutionarily related with the gene families of natural killer cell-specific receptors. *Eur J Immunol* **24**(7), 1692-7.

SHIMIZU, Y., GUIDOTTI, L. G., FOWLER, P., and CHISARI, F. V. (1998). Dendritic cell immunization breaks cytotoxic T lymphocyte tolerance in hepatitis B virus transgenic mice. *J Immunol* **161**(8), 4520.

SIMMONS, G., REEVES, J. D., GROGAN, C. C., VANDENBERGHE, L. H., BARIBAUD, F., WHITBECK, J. C., BURKE, E., BUCHMEIER, M. J., SOILLEUX, E. J., RILEY, J. L., DOMS, R. W., BATES, P., and POHLMANN, S. (2003). DC-SIGN and DC-SIGNR bind ebola glycoproteins and enhance infection of macrophages and endothelial cells. *Virology* **305**(1), 115-23.

SMALL, M., and KRAAL, G. (2003). In vitro evidence for participation of DEC-205 expressed by thymic cortical epithelial cells in clearance of apoptotic thymocytes. *Int Immunol* **15**(2), 197-203.

SOILLEUX, E. J., BARTEN, R., and TROWSDALE, J. (2000). DC-SIGN; a related gene, DC-SIGNR; and CD23 form a cluster on 19p13. *J Immunol* **165**(6), 2937-42.

SUZUKI, T., TAKAGI, T., FURUKOHRI, T., KAWAMURA, K., and NAKAUCHI, M. (1990). A calcium-dependent galactose-binding lectin from the tunicate Polyandrocarpa misakiensis. Isolation, characterization, and amino acid sequence. *J Biol Chem* **265**(3), 1274-81.

TASSANEETRITHEP, B., BURGESS, T. H., GRANELLI-PIPERNO, A., TRUMPFHELLER, C., FINKE, J., SUN, W., ELLER, M. A., PATTANAPANYASAT, K., SARASOMBATH, S., BIRX, D. L., STEINMAN, R. M., SCHLESINGER, S., and MAROVICH, M. A. (2003). DC-SIGN (CD209) mediates dengue virus infection of human dendritic cells. *J Exp Med* **197**(7), 823-9.

TAYLOR, P. R., BROWN, G. D., REID, D. M., WILLMENT, J. A., MARTINEZ-POMARES, L., GORDON, S., and WONG, S. Y. (2002). The beta-glucan receptor, dectin-1, is predominantly expressed on the surface of cells of the monocyte/macrophage and neutrophil lineages. *J Immunol* **169**(7), 3876-82.

TESTI, R., PHILLIPS, J. H., and LANIER, L. L. (1989). T cell activation via Leu-23 (CD69). *J Immunol* **143**(4), 1123-8.

TESTI, R., PULCINELLI, F., FRATI, L., GAZZANIGA, P. P., and SANTONI, A. (1990). CD69 is expressed on platelets and mediates platelet activation and aggregation. *J Exp Med* **172**(3), 701-7.

TESTI, R., PULCINELLI, F. M., CIFONE, M. G., BOTTI, D., DEL GROSSO, E., RIONDINO, S., FRATI, L., GAZZANIGA, P. P., and SANTONI, A. (1992). Preferential involvement of a phospholipase A2-dependent pathway in CD69-mediated platelet activation. *J Immunol* **148**(9), 2867-71.

TRUMPFHELLER, C., PARK, C. G., FINKE, J., STEINMAN, R. M., and GRANELLI-PIPERNO, A. (2003). Cell type-dependent retention and transmission of HIV-1 by DC-SIGN. *Int Immunol* **15**(2), 289-98.

VIVIER, E., and DAERON, M. (1997). Immunoreceptor tyrosine-based inhibition motifs. *Immunol Today* **18**(6), 286-91.

WEIS, W. I., DRICKAMER, K., and HENDRICKSON, W. A. (1992). Structure of a C-type mannose-binding protein complexed with an oligosaccharide. *Nature* **360**(6400), 127-34.

WEIS, W. I., TAYLOR, M. E., and DRICKAMER, K. (1998). The C-type lectin superfamily in the immune system. *Immunol Rev* **163**, 19-34.

WILLMENT, J. A., LIN, H. H., REID, D. M., TAYLOR, P. R., WILLIAMS, D. L., WONG, S. Y., GORDON, S., and BROWN, G. D. (2003). Dectin-1 expression and function are enhanced on alternatively activated and GM-CSF-treated

macrophages and are negatively regulated by IL-10, dexamethasone, and lipopolysaccharide. *J Immunol* **171**(9), 4569-73.

WILSON, J. L., HEFFLER, L. C., CHARO, J., SCHEYNIUS, A., BEJARANO, M. T., and LJUNGGREN, H. G. (1999). Targeting of human dendritic cells by autologous NK cells. *J Immunol* **163**(12), 6365-70.

YIM, D., SOTIRIADIS, J., KIM, K. S., SHIN, S. C., JIE, H. B., ROTHSCHILD, M. F., and KIM, Y. B. (2002). Molecular cloning, expression pattern and chromosomal mapping of pig CD69. *Immunogenetics* **54**(4), 276-81.

YOKOYAMA, W. M., and PLOUGASTEL, B. F. (2003). Immune functions encoded by the natural killer gene complex. *Nat Rev Immunol* **3**(4), 304-16.

YOKOYAMA, W. M., RYAN, J. C., HUNTER, J. J., SMITH, H. R., STARK, M., and SEAMAN, W. E. (1991). cDNA cloning of mouse NKR-P1 and genetic linkage with LY-49. Identification of a natural killer cell gene complex on mouse chromosome 6. *J Immunol* **147**(9), 3229-36.

ZELENSKY, A. N., and GREADY, J. E. (2003). Comparative analysis of structural properties of the C-type-lectin-like domain (CTLD). *Proteins* **52**(3), 466-77.

Zhang W, Trible RP, and Samelson LE (1998) LAT palmitoylation: its essential role in membrane microdomain targeting and tyrosine phosphorylation during T cell activation. *Immunity* 9:239-46.

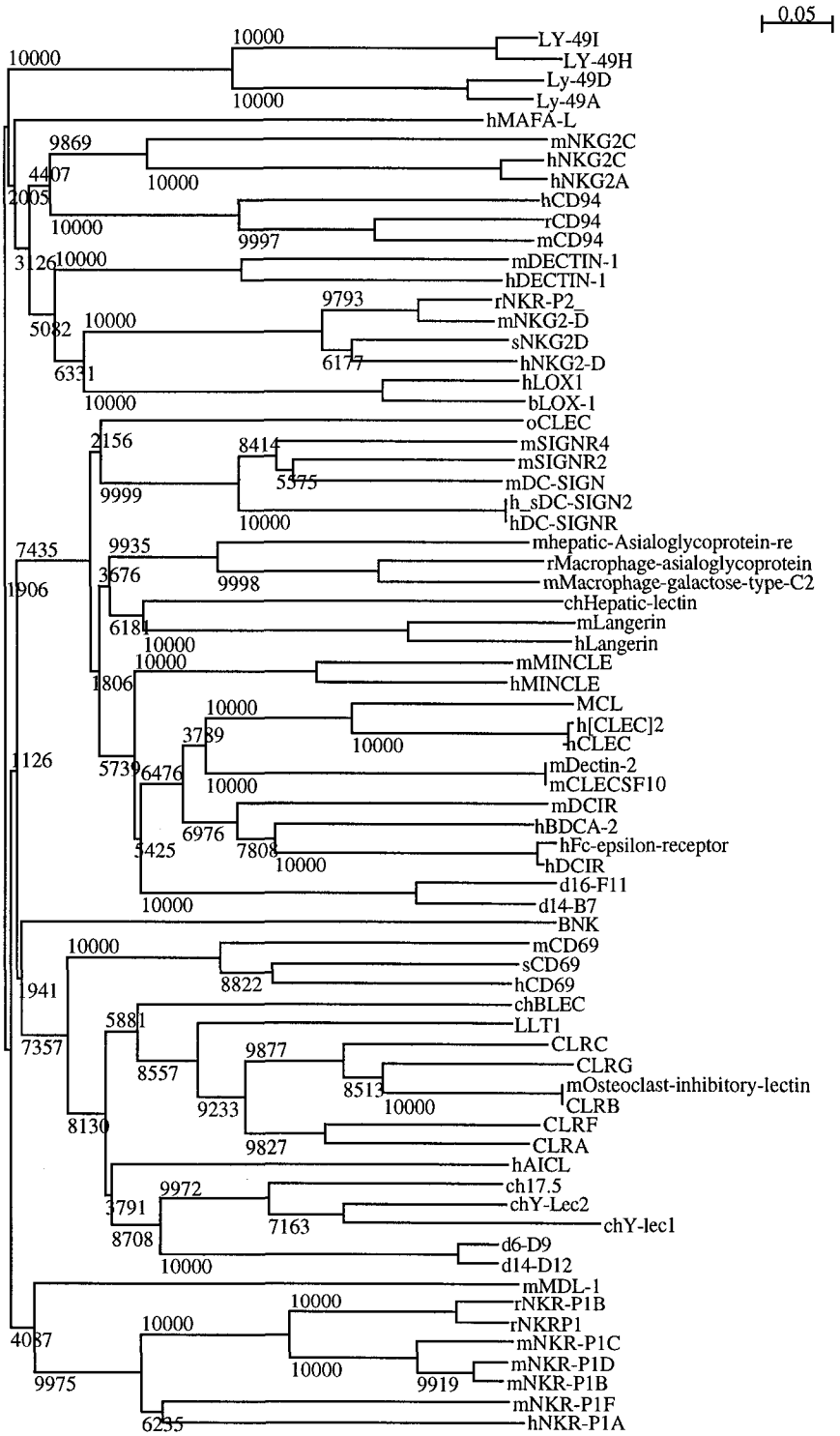
ZHOU, H., KARTSOGIANNIS, V., HU, Y. S., ELLIOTT, J., QUINN, J. M., MCKINSTRY, W. J., GILLESPIE, M. T., and NG, K. W. (2001). A novel osteoblast-derived C-type lectin that inhibits osteoclast formation. *J Biol Chem* **276**(18), 14916-23.

ZHOU, H., KARTSOGIANNIS, V., QUINN, J. M., LY, C., GANGE, C., ELLIOTT, J., NG, K. W., and GILLESPIE, M. T. (2002). Osteoclast inhibitory lectin, a family of new osteoclast inhibitors. *J Biol Chem* **277**(50), 48808-15.

ZIEGLER, S. F., LEVIN, S. D., JOHNSON, L., COPELAND, N. G., GILBERT, D. J., JENKINS, N. A., BAKER, E., SUTHERLAND, G. R., FELDHAUS, A. L., and RAMSDELL, F. (1994). The mouse CD69 gene. Structure, expression, and mapping to the NK gene complex. *J Immunol* **152**(3), 1228-36.

ZINGONI, A., PALMIERI, G., MORRONE, S., CARRETERO, M., LOPEZ-BOTEL, M., PICCOLI, M., FRATI, L., and SANTONI, A. (2000). CD69-triggered ERK activation and functions are negatively regulated by CD94 / NKG2-A inhibitory receptor. *Eur J Immunol* **30**(2), 644-51.

Appendices:



Appendix 1: Dendrogram based on an alignment of the CRD sequences of 86 lectins and lectin-like receptors with type II transmembrane orientation, including the four unique duck proteins.

```
   >14B7      GTGCTTCCAGTGTTCAGCATGGATGCCTTTGACAGCCAAAGGATGAATGAG 51
>16F11RT  GGAC.G.AC..TT.-..GG..A.GT....-..G..G....G.T.--.....CC.. 51

   >14B7  CAATAGAAAAGGTTCAGCAACTCAGTGGTGTCTTCCTCTGGATCCAGGGACCTAA 106
>16F11RT  ...G...G.--.-..-.G----..-C...-.A..G.----A.CC.........AG. 89

   >14B7 AGGGCCAAGCTTTTATATCACAAAACCAAAAACGGGGATTGGTGAGCACACTGTG 161
>16F11RT ....AGC....G..----------..-----..------CC.....-.C...G. 121

   >14B7 ATTAAATGGAAGCAGAAATCACCTATGCCGAAGTGAAGTTCAAGAATG-CATCAC 215
>16F11RT TC.TCC.C.CTT.T.----.-...-.....----------....A.C..C.C...- 160

   >14B7 CAACTGAAGAGGTTGAAGTACCTCAGAAGAAGCAGCAGCAGCATGAGCAACATAC 270
>16F11RT TG..C.TT.GCC.C.TT..-T...--TTTC.-...G...T.--..-..C.-G... 207

   >14B7 GCAGAC-ATGCCCTCCATGGCTCCCGTGGCTGATCTCACTGCTCCTGCTCCTGGT 324
>16F11RT -A....TC.....CAG.AT.....A.-..-..---G.....-----A...---- 247

   >14B7 GTGCGTTGCCCTTGTTGTTGTTCTCCTAGTCACTCACGTCCCCCAGAGCTGTGAC 379
>16F11RT --------..AA..GATC..CAAG..A.A.......T...T.............. 294

   >14B7 AAGCCCGCAGTCCTGCAGCGGAACCACACAGGGTGGCACTGCATCTTGGCAGTGC 434
>16F11RT ..........C.......G.......T....A..T.............CG.... 349

   >14B7 ATCAAGGCAAAGAG-GACAGCTGGAAGTGCTGTCCAGAGGGCTGGAGACCCTTTC 488
>16F11RT ............G.C..G.C..-.........T..................... 403

   >14B7 AGGAAAGCTGCTATTACTTCTCAGATGATCAGATGCCCTGGAATGAGAGCAAGAA 543
>16F11RT .........................................G........C..C. 458

   >14B7 GAACTGCAGTGGGATGGGCTCCCAGCTGGTGGTGATCAATACAGAAGCAGAGCAG 598
>16F11RT .............................................A.......... 513

  >14B7  GATTTCCTCTATAAGGAAATAAGAAGACAGATGAAATACCAACAAAATGCAATCA 653
>16F11RT .C.....................------.............G........G..... 562

   >14B7 ATTTATTCATCGGTCTGAGGGCACAGGAGGTGGGCCAGTGGCGCTGGGCAGACCA 708
>16F11RT ......A..................A............................ 617

   >14B7 GACTCCCTATAATGAATCAGCAGCGTTCTGGAGGTCTGGGGAGCCAAGTAATAAA 763
>16F11RT ...............AG..............CG............G..C.. 672

   >14B7 TC---TGATGAGCTGTGTGTTGTAATCCATCACAAAACAGAAAACCTCCGGAACT 815
>16F11RT C.AAG.........................T..C.G.A...T.TTT......... 727

   >14B7 GGAATGATGTCCCGTGCAGAATACGTTCTTATCGGATTTGTGAGACTGCAGCAGT 870
>16F11RT .....A.......A....C...C.AC......T....................A 782

   >14B7 AACTCTATGATGGAGGAATCCTCATCCTGAGATTAGCAGCGAACTGGGAACAGCA 925
>16F11RT ...AA.............................G..................... 837

   >14B7 GAGGGCTGTGTTGGGAGGGGTGGGAGAGCCTTGGAGCCTTTATCTTGCCTCTGCT 980
>16F11RT ....................A..............A...G.............. 892


   >14B7 GGTGGGATGATGAGACTGGGAGTGATGTTGCTCTGCACACAGCATCCCCTGTGCA 1035
>16F11RT C.............................................T...... 947
```

```
   >14B7 TGTGTATTTCTCAAAGAACTATGACGCCTCAAGTAGAAACAATAAATGCTAGAGA 1090
>16F11RT ....................-----.-..C...G.T..G..................... 996

   >14B7 ACTCTGAAAAAAAAAAAAAAAAAAAA                                  1117
>16F11RT ........................AAAAAAA                              1030
```

Appendix 2: Nucleotide alignment of the apLec-1 and the corrected form of apLec-2. Dots indicate identical residues, dashes indicate gaps introduced to maximize the alignment

```
 >29D1 GCAGCACAATCAAGACATGAGACACCAGGAGTAGACACCCTGCTCTTAGGGCACTC  56


 >29D1 CCACTACTTCAGTTTTCATCAGGCTGAGTGCTCCACCTCCCACACTGTTACTGC-G 111
 >33E5                                              ...-A.T...A.  11

>14D12                                        AGGCAGAGTGCTGCTGGGCAT  21
 >22G6                                        .....................  21
 >29D1 CACAAACACGATGACTTACATAAGACTCCAGGACT..... 167
 >33E5 .....T..A.-------....G....A......G...A..CCC....CT.A.....  60

  >6D9         GTTGCTGTGAGAGGACTGGATGTGTGGATGTGATCAGTGAGACCCGA  47
>14D12 GGTCAGCCA..... 77
 >22G6 GGTCAGCCA..... 77
 >29D1 GGTCAGCCA..... 223
 >33E5 GGTCAGCCA..... 116

  >6D9 CATTTCCTTGCCAGCCATGGGGAAAGGAGCCCAAGAAAAAAATCACCCTGACCAAG 103
>14D12 .....C..... 133
 >22G6 .....C..... 133
 >29D1 .....C..... 279
 >33E5 .....C..... 172

  >6D9 AAGAAGTAATGAACCTTCCAAGAGATGAAGAGAAACAATGCAAATGGGGCTTCAGC 159
>14D12 ........C......C.....G..A.....A..C.... 189
 >22G6 ........C......C.....G..A.....A..C.... 189
 >29D1 ........C......C.....G..A.....A..C.... 335
 >33E5 ........C......C.....G..A.....A..C.... 228

  >6D9 CCCCATGGGATGAAAAAGAAATGTCGTCGTGTAAAAAAGCTCCTGACTCCGCTGTG 215
>14D12 ........A...G.....C..C........G..... 245
 >22G6 ........A...G.....C..C........G..... 245
 >29D1 ........A...G.....C..C........G..... 391
 >33E5 ........A...G.....C..C........G..... 284

  >6D9 TGTGGTGCTGACTGTCCTTGTCCTCGCTCTGGTGGTGGCCTTGGTTGTTGTGCTTC 271
>14D12 ..........G....A.....C.....A.. 301
 >22G6 ..........G....A.....C.....A.. 301
 >29D1 ..........G....A.....C.....A.. 447
 >33E5 ..........G....A.....C.....A.. 340

  >6D9 TGCAGTCTCACTCATCACATCCCCAATTCTCCGACGTGTGCCCAGACAAATGGATC 327
>14D12 .....C.....CC..... 357
 >22G6 .....C.....CC..... 357
 >29D1 .....C.....CC..... 503
 >33E5 .....C.....CC..... 396

  >6D9 GGCTTCCAAAGCAAGTGCTACTATTTCTCGGAGGATGAAAGCAACTGGAAAACCAG 383
>14D12 .....TA..A..... 413
 >22G6 .....TA..A..... 413
 >29D1 .....TA..A..... 559
 >33E5 .....TA..A..... 452
```

```
  >6D9 CTTGGAGAACTGCAAGGCCATGGAAGCCTCCCTGACCTCCATAGACAGCCAGGAGG 439
>14D12 ................A..................................... 469
 >22G6 ................A..................................... 469
 >29D1 ................A..................................... 615
 >33E5 ................A..................................... 508

  >6D9 AACTGGCTTTCATCAAGCGCTTCAAGGGCCAAGCAAACCACTGGTTCGGGCTGCAC 495
>14D12 ......A.............A......A.......................... 525
 >22G6 ......A.............A......A.......................... 525
 >29D1 ......A.............A......A.......................... 671
 >33E5 ......A.............A......A.......................... 564

  >6D9 GACGAAGACAACAGCCAGTGGAGGTGGACCAACGGCGCAGCCTTCAACAACTGGTT 551
>14D12 .......G.................................................. 581
 >22G6 .......G.................................................. 581
 >29D1 .......G.................................................. 727
 >33E5 .......GC................................................. 620

  >6D9 TGAGGTGCGGGGAGGTGGCCCTTGTGCGTACATAAACCAGG---AGATCAGCTCAG 604
>14D12 ..........................................AGA............ 637
 >22G6 ..........................................AGA............ 637
 >29D1 ..........................................AGA............ 783
 >33E5 ..........................................AGA............ 676

  >6D9 CCTTCTGCAACACGGAGAAATACTGGATCTGCAGCAGGCCCAACAACTACGTCCTC 660
>14D12 ......................................................... 693
 >22G6 ......................................................... 693
 >29D1 ......................................................... 839
 >33E5 ......................................................... 732

  >6D9 TGGAGGCAAAAGATTTACCCCGAAT-AAAGATCTTATAAACATAAAAAAAAAAAAA 710
>14D12 ..................TT.G.A......GCC............CTCATGATA 749
 >22G6 ..................TT.G.A......GCC............CTCATGATA 749
 >29D1 ..................TT.G.A......GCC............CTCATGATA 895
 >33E5 ..................TT.G.A......GCC............CTCATGATA 788

  >6D9 AAAAAA                                                721
>14D12 GTGACAGAAAAAATGGTGATTCTTATTTAGGTCTGCTTCTTTGCCACAACAGTAGT 805
 >22G6 GTGACA..................................................... 805
 >29D1 GTGACA..................................................... 951
 >33E5 GTGACA..................................................... 844

>14D12 GCGGAGTTTCCTTTCTGCTCAGATTTGTGCAGACTCTGTCAATCTTTTGTACTGAA 861
 >22G6 ......................................................... 861
 >29D1 ......................................................... 1007
 >33E5 ......................................................... 900

>14D12 GTAAGCGACATTCCTTTTTCTTTCTCTGCTTCTACAGCAGAAAACAGTGTTTTATT 917
 >22G6 ......................................................... 917
 >29D1 ..................................................T..... 1063
 >33E5 ......................................................... 956

>14D12 TTTTTCTACATTTGCGAAAGTTATGTTTGAAATAATGAATTTGAAACCTGTGGATC 973
 >22G6 ......................................................... 973
 >29D1 ......................................................... 1119
 >33E5 ......................................................... 1012
```

```
>14D12 TGAAGTGCTATGGAATTATGCTATTATGCTATTATACTTCACACTACCGGTTGGAA 1029
 >22G6 ..........................------------------------- 1003
 >29D1 ..................................................... 1175
 >33E5 ..................................................... 1068

>14D12 GGAGGATAAGGAATGATTGGAACAGATGGACTTATGCTATTATGCTGTTATACTTC 1085
 >22G6 --------------------.----------------------.-..-..... 1013
 >29D1 ..................................................... 1231
 >33E5 ..................................................... 1124

>14D12 ACTGTACATATTGGAAGGATGAGAAGGAACGATTGGAACAGATAATGTCTGGATGT 1141
 >22G6 ..................................................... 1069
 >29D1 ..................................................... 1287
 >33E5 ..................................................... 1180

>14D12 GTGTATCTGGATGGATTTTTTAATAAAAAAATAAAAATTGGAAGCTAATTGAAGCA 1197
 >22G6 ..................................................... 1125
 >29D1 ..................................................... 1343
 >33E5 .....................................................-------.. 1228

>14D12 GAGCCTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA               1239
 >22G6 ..----.....................                          1148
 >29D1 .......C..........................                  1378
 >33E5 ------.............................                  1259
```

Appendix 3: Nucleotide alignment of the five duck AICL-like cDNAs. Dots indicate identical residues, dashes indicate gaps introduced to maximize the alignment

```
>D105 draft          CACCCTGACCAAGAAGAAGTACTGAACCCTCCAAGATATGAAT 43
     >132-1          CACCCTGACCAAGAAGAAGTACTGAACCCTCCA▒GAGATGA▒G 43
     >132-2          CACCCTGACCAAGAAGAAGTACTGAACCCTCCAAGA▒ATGA▒G 43
 >D64 draft          CACCCTGACCAAGAAGAAGTAATGAACCTTCCAAGAGATGAAG 43
       >129          CACCCTGACCAAGAAGAAGTACTGAACCCTCCAAGAGATGA▒G 43
      >95-2 GAAAAAAATCACCCTGACCAAGAAGAAGTAATGAACCTTCCAAGAGATGAAG 52
      >95-1          CACCCTGACCAAGAAGAAGTACTGAACCCTCCAAGAGATGAGG 43

>D105 draft AAAAGCAATATAAATGGGGTAATGCTGCGGGAAAGCGAGATCTGGCTCCAGT 95
     >132-1 AAAAACAATGCAAATGGGGTAATGCTGGGGGAAAG▒GAGATCTGGCTCCAGT 95
     >132-2 AAAA▒CAAT▒▒AAATGGGGTAATGCTG▒GGGAAAG▒GAGATCTGGCTCCAGT 95
 >D64 draft AGAAACAATGCAAATGGGGTRATGCTGGGGGAAAGCGAGATCTGGCTCCAGT 95
       >129 AAAAACAATGCAAATGGGGTAATGCTGGGGGAAAG▒GAGATCTGGCTCCAGT 95
      >95-2 AGAAACAATGCAAATGGGGTAATGCTGGGGGAAAGCGAGATCTGGCTCCAGT 104
      >95-1 AAAAACAATGCAAATGGGGTAATGCTGGGGGAAAGTGAGATCTGGCTCCAGT 95

>D105 draft GGGGCACTGGTGGGATGGGCACCAGCCAGGG-TGTTTTGGTGGGTGAGGACA 146
     >132-1 GGGGCACTGGTGGGATGGGCACCAGCCAGGG-TGTTTTGGTGGGTGAGGACA 146
     >132-2 GGGGCACTGGTGGGATGGGCACCAGCCAGGG-T▒TTTTGGTGGGTGAGGACA 146
 >D64 draft GGGGCACTGGTGGGATGGGCACCAGCCAGGG-TGTTTTGGTGGGTGAGGACA 146
       >129 GGGGCACTGGTGGGATGGGCACCAGCCAGGG-TGTTTTGGTGGGTGAGGACA 146
      >95-2 GGGGCACTGGTGGGATGGGCACCAGCCAGGG-TGTTTTGGTGGGTGAGGACA 155
      >95-1 GGGGCACTGGTGGGATGGGCACCANCCAGGGNTGTTTTGGTGGGTGAGGACA 147

>D105 draft GACTCTGAATTTTACCATGTAGCCACCTCTTAATNCACCTGGAAGACCTCAA 198
    >D132-1 GACTCTGAATTTTA▒CATG▒▒▒CCACCTCTTAAT-CACCTGGAAGACCTCAA 197
     >132-2 G▒CTCTGAATTTTA▒CATG▒▒▒CCACCTCTTAAT-CACCTGGAAGACCTCAA 197
  >D64 draft GACTCTGAATTTTATCATGTAGCCACCTCTTAATCCACCTGGAGGACCTCAA 198
       >D129 GACTCTGAATTTTA▒CATG▒▒▒CCACCTCTTAAT-CACCTGGAAGACCTCAA 197
      >95-2 GACTCTGAATTTTATCATGTAGCCACCTCTTAATCCACCTGGAGGACCTCAA 207
      >95-1 GACTCTGAATTTTANCATGCTACCACCTCTTAAT-CACCTGGAANACCTCAA 198

>D105 draft AGAAGGGAACTCCTATCCCAAGGGGGAGGGCTATAGGATCAACCATGGATCC 250
    >D132-1 AGAAGGGAACTCCTATCCCAAGGGGGAGGG▒TATAGGATCAACCATGGATCC 249
     >132-2 AGAAGGGAACTCCTATCCCAAGGGGGAGGG▒TATAGGATCAACCATGGATCC 249
 >D64 draft AGAAGGGAGCTCCTATCCCAAGGGGGAGGGCTATAGGATCAACCGTGGATCC 250
      >D129 AGAAGGGAACTCCTATCCCAAGGGGGAGGG▒TATAGGATCAACCATGGATCC 249
      >95-2 AGAAGGGAGCTCCTATCCCAAGGGGGAGGGCTATAGGATCAACCGTGGATCC 259
      >95-1 ANAAGGGAACTCCTATCCCAAGGGRGAGGGTTATAGGATCAACCATGGATCC 250

>D105 draft AGCAGAAAAAGGGCAAATGGGGGGTGGAAAGACCCAGCCCACATCTCTTGCT 302
    >D132-1 AGCAGAAAAAGGGCAAATGGGGGGTGGAAAGACCCAGCCCACATCTCCTGCT 301
     >132-2 AGCAGAAAAAGGGCAAATGGGGGGTGGAAAGACCCAGCCCACATCTC▒TGCT 301
 >D64 draft AGCAGAAGAAGGGCAAATGGGGGGTGGAAAGRNCCAGCCCACATCTCCTGCT 302
      >D129 AGCAGAAAAAGGGCAAATGGGGGGTGGAAAGACCCAGCCCACATCTCCTGCT 301
      >95-2 AGCAGAAGAAGGGCAAATGGGGGGTGGAAAGACCCAGCCCACATCTCCTGCT 311
      >95-1 AGCANAAAAAGGGCAAATGGGGGGTGGAAAGACCCAGCCCACATCTCCTGCT 302

>D105 draft CTCTTCTCCACCRGACTCCAGCCCCCATGGGATGGAAAAGAAATGTCGTCGT 354
    >D132-1 CTCTTCT▒CACCAGACTCCAGCCCCCATGG▒ATGGAAAAGAAATGTCGTCGT 353
     >132-2 CTCTTCT▒CACCAGACTCCAGCCCCCATGGGATGGAAAAGAAATGTCGTCGT 353
 >D64 draft CTCTTCTCCACCAGGCTTCAGCCCCCATGGGATGAAAAAGAAATGTCGTCGT 354
      >D129 CTCTTCT▒CACCAGACTCCAGCCCCCATGGGATGGAAAAGAAATGTCGTCGT 353
      >95-2 CTCTTCTCCACCAGGCTTCAGCCCCCATGGGATGAAAAAGAAATGTCGTCGT 363
      >95-1 CTCTTCTACACCANACTCCAGCCCCCATGGAATGGAAAAGAAATGTCGTCGT 354
```

```
>D105 draft GTACAACAGCTCCTGGCTCCGCTATGTGTGGTGCTGAGTGTCCTTGTCCTCG 406
    >D132-1 GTACAACAGCTCCTGGCTCCGCTGTGTGTGGTGCTGAGTGTC▒TTGTCCTCG 405
     >132-2 GTACAACAGCTCCTGGCTCCGCT▒TGTGTGGTGCTGAGTGTCCTTGTCCTCG 405
 >D64 draft GTAAAAAAGCTCCTGACTCCGCTGTGTGTGGTGCTGASTGTCCTTGTCCTCG 406
      >D129 GTACAACAGCTCCTGGCTCCGCTGTGTGTGGTGCTGAGTGTCCTTGTCCTC▒ 405
      >95-2 GTAAAAAAGCTCCTGACTCCGCTGTGTGTGGTGCTGACTGTCCTTGTCCTCG 415
      >95-1 GTACAACAGCTCCTGGCTCCGCTGTGTGTGGTGCTGAGTGTCATTGTCCTCG 406

>D105 draft CTCTGCTGGTGGNCTTGGTTGGTGAGTACC-ACTGGAGCATCCCCACGGCAG 457
    >D132-1 CTCTGCTGGTGGCCTTGGTTGGTGAGT▒CC-ACTGGAGCATCCCCACGGCAG 456
     >132-2 CTCTGCTGGTGGCCTTGGTTGGTGAGT▒CC-ACTGGAGCATCCCCACGGCAG 456
 >D64 draft CTCTGSTGGTGGCCTTGGTTGGTGAGTSCC-AMTGGAGYRTNCCCACGGCAN 457
 >D129 CTCTGCTGGTGGCCTTGGTTGGTGAGT▒CC-ACTGGA▒CATCCCCAC▒GCAG 456
      >95-2 CTCTGGTGGTGGCCTTGGTTGGTGAGTCCC-ACTGGAGTGTCCCCACGGCAG 466
      >95-1 CTCTGCTGGTGGCCTTGGTTGGTGAGTACCNACTGGAGCATCCCCACGGYAG 458

>D105 draft GGA-GATGGGCCARGAACCTTTTYCAGGGATGGTGTGGGGGATTTCAGGGTT 508
    >D132-1 GGA-GATGGGCCAGGAAC▒TTTTTCAGGGATGGTGTGGGGGATTTCAGGGTT 507
     >132-2 GGA-GATGGGCCAGGAAC▒TTTTTCAGGGATGGTGTGGGGGATTTCAGGGTT 507
 >D64 draft GGA-GATNGGCCAGGAACNTTTTTCAGGGATGGTGTGGGGGATTTCAGGNTT 508
  >D129 GGA-GAT▒GGCCAGG▒AC▒T▒TTTCAGGGATGG▒GTGGGG▒ATT▒CAGGGTT 507
      >95-2 GGA-GATGGGCCAGGAACCTGTTTCAGGGATGGAGTGGGGAATTCCAGGGTT 517
      >95-1 GGANGATGGGCCAGGAACTTTTTTCAGGGATGGTGTGGGGGATTTCAGGGTT 510

>D105 draft TTCCCANCAGGATGAAAGGSCCCGAACCATAGCACTGAGCCTGCTTCATGGG 560
    >D132-1 TTCCCAGCAGGATGAAAGGCCCCGAACCATAGCA▒TGAGCCTG▒TTCATGGG 559
     >132-2 TTCCCAGCAGGATGAAAGGCCCCGAACCATAGCACTGAGCCTGCTTCATGGG 559
 >D64 draft TTCCCAGCAGGATGAAAGGCCCCGAACCATAGCACTGAGCCTGCTTCATGGG 560
      >D129 T▒CCCAGCAGGATGA▒A▒GCCCC▒AA▒CATAGCACTGAGC▒TGCTTCATG▒▒ 559
      >95-2 TTCCCAGCAGGATGAGAAGCCCCCAAGCATAGCACTGAACCTGCTTCATGGT 569
      >95-1 TTCCCAGCAGGATGNAAGGCCCCGAACCATAGCANTGAGCCTGCTTCATGGG 562

>D105 draft GCACCCGCTTTTGTCCCCTCAANNNNNNNNNNNNNNNNNNNNNNNNNNNN 612
    >D132-1 GCACCC▒CTTTTGTCCCCTCAACCAGG▒T▒TA▒TCAGCCTTTCCAAGC▒GGT 611
     >132-2 GCACCC▒CTTTTGTC▒CCTCAACCAGG▒T▒TACTCAGCCTTTCCAAGC▒GGT 611
 >D64 draft GCACCCCNTTTAGTCCCCTCAACCAGGNTGTACTCAGCCTTTCCAAGCGGGT 612
      >D129 GCACCC▒CTTTTGTCCCC▒CAACCAGG▒T▒TACTCAGCCTTTCCAAGC▒GGT 611
      >95-2 GCACCCGCTTTTGTCCCCTCAACCAGGCTTTACTCAGCCTTTCCAAGCGGGT 621
      >95-1 GCACCCCCTTTTGTCCCCTCAACCAGGCTYTACTCAGCCTTTCCAAGCANGT 614

>D105 draft NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-NNNNNNNNNNNNNNN 663
    >D132-1 GCCCC▒T▒TTGG▒GCATCCCAACCCCCCTAC▒▒TT-G▒TGGAGGGGAACACA 662
     >132-2 GCCCC▒T▒TTGG▒GCATCCCAACCCCCCTAC▒▒TT-G▒TGGAGGGGAACACA 662
 >D64 draft GCCCCTTTTTGGCGCATCCCAACCCCCATACGCTT-GCTGGAGGGGAACACA 663
      >D129 GCCC▒▒T▒TTG▒▒GCATCCCAA▒CCCCCTAC▒▒TT-G▒TGGAGGGGAACACA 662
      >95-2 GCCCCTTCTTGGCGCATCCCAACCCCCCTACATTT-GCTGGAGGGGAATACA 672
      >95-1 GCCCCATTTTGGNGCATCCNAACCCCCCTACGCTTNGNTGGAGGGGAACACA 666

>D105 draft NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 715
    >D132-1 TATAATTAA▒▒AAATGTTATTTTT▒TGC▒TTTTTT▒▒--▒▒TTTTTTTTTT▒ 710
     >132-2 TATAATTAAA▒AAATGTTATTTTT▒TGC▒TTTTTT▒▒--▒▒TTTTTTTTTT▒ 708
 >D64 draft TATAACTAAAAAAATGTTATTTTTTTGC----------CNTTTTTTTTTTT 704
      >D129 TATAATTAAA▒AAA▒▒▒▒▒▒▒▒▒▒▒▒GC▒TTTTTT▒▒▒▒▒▒TTTTTTTTTT▒ 703
      >95-2 TATAATTTAAAAAATGTTATTTTTCTGCTTTTTTTTTTTTTTTTTTTTTTT 724
      >95-1 TATAATTAAAGAAATGTTATTTTTTTGCCTTTTTTT-----TTTTTTTTTC 713
```

```
>D105 draft NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 767
>D132-1 [redacted]GTGGCTGTCCATGCATCTCACCTG[redacted]TTGGTTTTCTCCTTTTACAGTTGT 762
    >132-2 [redacted]GTGGCTGTCCATGCATCTCACCTG[redacted]TTGGTTTTCTCCTTTTACAGTTGT 760
>D64 draft TTGGTGGCTGTCCATGCATCTCACCTGCTTGGTTTTCTCCTTTGACAGTTGT 756
>D129 [redacted]GTGGCTGTCCATGCATCTCACCTG[redacted]TTGGTTTTCTCCTTTTACAGTTGT 755
     >95-2 TTTGTGGCTGTCCGTGCATCTCACCTGATCGGTTTTCTCCTTTTACAGTTGT 776
     >95-1 CNNGTGGCTGTCCATGCATCTCACCTGNTTGGTTKTCTCCTTTTACAGTTGT 765

>D105 draft NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCC 819
    >D132-1 GCA[redacted]TGCAGTCTCACTCATCACATCCCC[redacted]ATTCTCCCACGTGTGCCCA[redacted]C 814
     >132-2 GCA[redacted]TGCAGTCTCACTCATCACATCCCC[redacted]ATTCTCCCACGTGTGCCCA[redacted] 811
 >D64 draft GCAGCTGCAGTCTCGCTCATCACATCCCCAATTCTCCCACGTGTGTCCAGNN 808
      >D129 GCA[redacted]GCAGTCTCACTCATCACATCCCC[redacted]ATTCTCCCACGTGTGCCCA[redacted]C 807
      >95-2 GCTTCTGCAGTCTCACTCATCACATCCCCAATTCTCCGACGTGTGCCCAGAC 828
      >95-1 GCATNTGCRGTCTCACTCATCACATCCCCCATTCTCCCANGTGTGCCCMCMC 817

>D105 draft -AAATGGATCGGCTT-CCAAAGCAAGTGCTACTATTTCTCNRNGA-ATGAAA 868
    >D132-1 -AAATGGATCGGCTT[redacted]CC[redacted]A-ATGAAA 864
     >132-2 -[redacted]A-ATGAAA 861
 >D64 draft -NNNNNNNNNNNNNNNNNNNNNNNNNTGCTACTATTTCTCAGAGG-ATGAAA 858
      >D129 -AAATGGATCGGCTT-CC[redacted]TGC[redacted]ACTAT[redacted]TCTC[redacted]A-ATGAAA 856
      >95-2 -AAATGGATCGGCTT-CCAAAGCAAGTGCTACTATTTCTCGGAGG-ATGAAA 877
   >95(6D9)                                                GATGAAA 7
      >95-1 NAAATGGATCGGCTT-CCAAAGCAAGTGCNACTATKTCTCTAANANATGAAA 868

>D105 draft GCAACTGGAAAACCAGCTTGGAGAACTGCAAGACCATGGAAGCCTCCCTGAC 920
    >D132-1 GCAACTGGAAAACCAGCTTGGAGAACTGCAAGACCATGGAAGCCTCCCTGAC 916
     >132-2 GCAACTGGAAAACCA[redacted]CTTGGAGAAC[redacted]G[redacted]GACCATGGAAGCCTCCCTGAC 913
 >D64 draft GCAACTGGAAAACCAGCTTGGAGAAGTGCAAGGCCATGGAAGCCTCCCTGAC 910
      >D129 GCAACTGGAAAACCAGCTTGGAGAACTGCAAGACCATGGAAGCCTCCCTGAC 908
      >95-2 GCAACTGGAAAACCAGCTTGGAGAACTGCAAGGCCATGGAAGCCTCCCTGAC 929
 >D95 (6D9) GCAACTGGAAAACCAGCTTGGAGAACTGCAAGGCCATGGAAGCCTCCCTGAC 59
      >95-1 GCNACTGGAAAACCANCTTGGAGAACTGNNAGACCATGGAAGCCTCCCTGAC 920

>D105 draft CTCCATAGACAGCCAGGAGGAACTGGTGAGAACTNNNNNNGAGCCCCAACAC 972
    >D132-1 CTCCATAGACAGCCAGGAGGAACTGGTGAGAACT[redacted]GAGCCCC[redacted]ACAC 968
     >132-2 CTCCATA[redacted]ACAGCCAGGAGGAACTGGTGAGAACT[redacted]GA[redacted]CCCCAA[redacted]AC 965
 >D64 draft CTCCATAGACAGCCAGGAGGAACTGGTGAGAACTGGGCACGAGTCCCAACAC 962
      >D129 CTCCATAGACAGCCAGGAGGAACTGGTGAGAACT[redacted]GAGCCCCAA[redacted]C 960
      >95-2 CTCCATAGACAGCCAGGAGGAACTGGTGAGAACTGGGCACGAGCCCCAACAC 981
 >D95 (6D9) CTCCATAGACAGCCAGGAGGTACTGGTGAGAACTGGGCACGAGCCCCAACAC 111
      >95-1 CTCCATANACAGCCAGGAGGAACTGGTGAGAACT------GAGCCCCTACAC 966

>D105 draft AGGCTCTGGCGTGGAAAATGTGTGGGACTARAACAAAATCCAATTTTCCCCC 1024
    >D132-1 AGGCTCTGG[redacted]GTGGAAAATGTGTGGGACTA[redacted]A[redacted]CAAAATCCAATTTTCCCCC 1020
     >132-2 AGGCTC[redacted]GG[redacted]GTGGAAAATGTGTGGGACTA[redacted]A[redacted]CAAAATCCAATTTTCCCCC 1017
 >D64 draft AGGCTCTGGTGTGGAAAATGTGTGGGACTAGACCAAAATCCAATTTTCCCCC 1014
      >D129 [redacted]G[redacted]TCTGG[redacted]GTGGAAAATGTGTGGGACTA[redacted]A[redacted]CAAAATCC[redacted]ATTTTCCCCC 1012
      >95-2 AGGCTCTGGTGTGGAAAATGTGTGGGACTAGACCAAAATCCCATTTTCCCCC 1033
 >D95 (6D9) AGGCTCTGGTGTGGAAAATGTGTGGGACTAGACCAAAATCCCATTTTCCCCC 163
      >95-1 AGGCTCTGGNGTGGAAAATGTGTGGGACTARAACAAAATCCAATTTWCCCCC 1018
```

```
>D105 draft AAATTTCACCCACTGCCAAAGCYAGAANAAGATGCCCTGGGGAAAGCAACAG 1076
    >D132-1 AAATTTCACCCACTGCCAAAGCCA[N][N]A[N]AAGATGCCCTGGGGAAAGCAACAG 1072
     >132-2 AAATTTCACCCACTGCCAAAGCCA[N][N]A[N]AAGATGCCCTGGGGAAA[N]CAACAG 1069
 >D64 draft AAATTTCACCCACTGCCAAAGCCAGGAGAAGATGCCCTGGGGAAAGCAACAG 1066
      >D129 AAATTTCACCCACTGCCAAAGCCAG[N]A[N]AAGATGCCCTGGGGAAAGCAACAG 1064
      >95-2 AAATTTCACCCACTACCAAAGCCAGGAGAAGATGCCCTGGGGAAAGCAACAG 1085
 >D95 (6D9) AAATTTCACCCACTGCCAAAGCCAGGAGAAGATGCCCTGGGGAAAGCAACAG 215
      >95-1 AAATTYCACCCACTGCCAAAGCCANAANAANATGCCCTGGGGAAAGCAACAG 1070

>D105 draft CCCAGGCTGGGGAGTGAGCAATATCACATTATYTTCTCCAAAACTCTCTCCC 1128
    >D132-1 CC[N][N]GGCTG[N]GGAGTGAGCAAT[N]T[N]ACATTATTTTCTC[N]AAAACTC[N]CTCCC 1124
     >132-2 CC[N][N]GGCTGGGGAGTGAGCAAT[N]T[N]ACATTATTTTCTC[N]AAAACTC[N]CTCCC 1121
 >D64 draft CCTGGGCTGGGGAGTGTGCAATTTTACATTATTTTCTCAAAAACTCCCTCCC 1118
      >D129 CC[N][N]GGCTGGGGAGTGAGCAAT[N]T[N]ACATTATTTTCTC[N]AAAACTC[N]CTCCC 1116
      >95-2 CCTGGGCTGGGGAGTGAGCAATTTTACATTATTTTCTCAAAAACTCCCTCCC 1137
 >D95 (6D9) CCTGGGCTGGGGAGTGAGCAATTTTACATTATTTTCTCAAAAACTCCCTCCC 267
      >95-1 CCCAGGCTGTGGAGTGAGCAATATCACATTATTTTCTCCAAAACTCTCTCCC 1122

>D105 draft TATGTCCAGATGGGTGCTGTCCCGATCTTTGCCCTCAAGGGGCTTCCCTTTT 1180
    >D132-1 TA[N]GTCCAGATGGGTGCTGTCC[N]GA[N]CTTTGCCCTCAAGGGGCTTCC[N]TTTT 1176
     >132-2 TA[N]GTCCA[N]ATGGGTGCTGTCC[N]GA[N]CTTTGCCCTCAAGGGGCTTCC[N]TTTT 1173
 >D64 draft TACGTCCAGATGGGTGCTGTCCTGAGCTTTGCCCTCAAGGGGCTTCCTTTTT 1170
      >D129 TA[N]GTCCAGATGGGTGCTGTCC[N]GA[N]CTTTGCCCTCAAGGGGCTTCC[N]TTTT 1168
      >95-2 TACGTCCAGATGGGTGCTGTCCTGAGCTTTGTCCTCAAGGGGCTTCCTTTTT 1189
 >D95 (6D9) TACGTCCAGATGGGTGCTGTCCTGAGCTTTGTCCTCAAGGGGCTTCCTTTTT 319
      >95-1 TATGTCCANATGGGTGCTGTCCCGATCTTTGCCCTCAAGGGGCTTCCCTTTT 1174

>D105 draft CTGTGAAACTGAAGAATATTCACATCTCTCAAATCAAATCACCTT-CCTATT 1231
    >D132-1 CTGTGAAACTGAAGA[N]TATTCACATCTCTCAAATCAAATCACCTT-CCTATT 1227
     >132-2 CTGTGAAACTGAA[N]AATATTCACATCTCTCAAATCAAATCACCTT-CCTAT[N] 1224
 >D64 draft CTGTGAAACTGAAGAATATTCACATCTCTCAAATCAAATCACCTT-CCTATT 1221
      >D129 CTGTGAAACTGAAGA[N]TATTCACAT[N]TCTCAAATCAAATCACCTT-CCTATT 1219
      >95-2 CTGTGAAACTGAAGAATATTCACATATCTCAAATCAAATCACCTA-CCTATT 1240
 >D95 (6D9) CTGTGAAACTGAAGAATATTCACATATCTCAAATCAAATCACCTTNCCTATT 371
      >95-1 CTGTGAAACTGAANAATATTCACATCTCTCAAATCAAATCACCTTNCCTATT 1226

>D105 draft GCA-GGGTTCCTATTGCTGGCTGAGTTGTATTTTAGGAAGACA-GGATCACC 1281
    >D132-1 GCA-GGGTTCCTATTGCTGGCTGAGTTGTATTTTAGGAAGACA-GGATCACC 1277
     >132-2 G[N][N]-GGGTTCCTATTGCTGG[N]TGAGTTGTATTTTAGGAA[N]A[N]A[N]GGATC[N]CC 1275
 >D64 draft GCA-GGGTTCCTATTGCTGGCTGAGTTTTGTTTTAGGAAGACA-GGATCACC 1271
      >D129 GCA-GGGTTCCTATTGCTGGCTGAGTTGTATTTTAGGAAGACA-GGATC[N]CC 1269
      >95-2 GCA-GGGTTCCTATTGCTGGCTGAGTTGTATTTTAGGAAGACA-GGATCACC 1290
 >D95 (6D9) GCANGGGTTCCTATTGCTGGCTGAGTTGTATTTTAGGAAGACA-GGATCACC 422
      >95-1 GCANGGGTTCCTATTGCTGGCTGAGTTGTATTTTAGGAANACA-GGATCNCC 1277

>D105 draft TAAACCCNNTTTATTTTTTTTTTTTCTTTTTTCGAT-AAAGGATTTCATCAAG 1332
    >D132-1 TAAACCC[N][N]TTT[N]TTTTTTTTTTTTCTTTTT[N]CGAT-AAAGG[N]TTTCATCAAG 1328
     >132-2 TAAACCC[N][N]TT[N][N]TTTTTTTTTTCTTTTT[N]CGAT-AAAGG[N]TTTCATC[N]AG 1326
 >D64 draft TAAACCCTTTTTTTTTTTTCTTTNNNTTTNNNCGAT-AAAGGCTTTCATCAAG 1322
      >D129 TAAACCC[N][N]TTT[N]TTTTTTTTTTTCTTT[N][N][N]CGAT-AAAGG[N]TTTCATCAAG 1320
      >95-2 TAAACCC--TTTTTTTTTTTTTTTTCTTTTTTCGAT-AAAGGCTTTCATCAAG 1339
 >D95 (6D9) TAAACCCTTTTT-TTTTTTTTTNNCTTTTTTCGATNAAAGGCTTTCATCAAG 473
      >95-1 TAAACCC--TTTATTTTTTTTTTTTCTTTTTNCGATNAAAGGATTTCATCAAG 1327
```

```
>D105 draft CGCTACAAGGGACAAGCAAACCANTGGTTCGGGC-TGCACGACGAAGGCAAC 1383
    >D132-1 CGCT▒CAAGGG▒CAAGCAAACCACTGGTTCGGG▒-TGCACGAC▒AAG▒CAAC 1379
     >132-2 CGCT▒CAAGGG▒CAAGCAAAC▒ACTGGTTCGGGC-TGCACGAC▒AAG▒CAAC 1377
 >D64 draft CGTTTCAAGGSCCAAGCAAACCACTGGTTCGGGCNTGCACGAYGAAGRCMAC 1374
      >D129 CGCT▒CAAGGG▒CAAGCAAACCACTGGTTCGGGC-TGCACGACGAAG▒CAAC 1371
      >95-2 CGCTTCAAGGGCCAAGCAAACCACTGGTTCGGGC-TGCACGACGAAGACAAC 1390
 >D95 (6D9) CGCTTCAAGGGCCAAGCAAACCACTGGTTCGGGC-TGCACGACGAAGACAAC 524
      >95-1 CGCTACAAGGGACAAGCAAACCACTGGTTCGGGC-TGCNCGACNAARGCAAC 1378

>D105 draft -AG-CCAGTGGAGGTGGACC-AACGGCGCA-GCNTTCAACAACTGG-TCAGT 1430
    >D132-1 -AG-CCAGTGGAGGTGGA▒C-AACGGCGCA-▒CCTTCAACAACTGG-TCA▒T 1426
     >132-2 -AG-CCAGTGGAGGTGGACC-AACGGCGCA-▒CCTTCAACAACTGG-TC▒GT 1424
 >D64 draft NAGNCCAGTGGAGGTGGACCNAACGGCGCA-GCCTTCAACAAYTGG-TCAGT 1424
      >D129 -AG-CCAGTG▒AGGTGGACC-AACGGCGCA-GCCTTCAACAA▒TGG-TCAGT 1418
      >95-2 -AG-CCAGTGGAGGTGGACC-AACGGCGCA-GCCTTCAACAACTGG-TCAGT 1437
 >D95 (6D9) -AG-CCAGTGGAGGTGGACC-AACGGCGCA-GCCTTCAACAACTGG-TCAGT 571
      >95-1 -AG-CCAGTGGAGGTGGACC-AACGGCGCANKMCTTCAACAACTGGNTCANT 1427

>D105 draft CCCTCCGCCTGTTGGGGTTGGGTGCTTTGGGTTGGCAATTAGGGCAGGAGTT 1482
    >D132-1 CCCTCCGCCTGTTGGGGTTGGGTGCTTTGGGTTGGCAATT▒GGGCAGGA▒TT 1478
     >132-2 CCCTCCGCCTGTTGGGGTTGGGTGC▒TT[redacted] 1475
 >D64 draft CCCTCCRCCTGTTGGGKTTGGGTGCTTTGGNTTGNCRATTAGGSCAGGAGTT 1476
      >D129 CCCTCCGCCTGTTGGGGTTGGGTGCTTTGGGTTGGCAATTA▒GGCAGGAGTT 1470
      >95-2 CCCTCCACCTGTTGGGGTTGGGTGCTTTGGGTTGGCAATTAGGGCAGGAGTT 1489
 >D95 (6D9) CCCTCCACCTGTTGGGGTTGGGTGCTTTGGGTTGGCAATTAGGGCAGGAGTT 623
      >95-1 CCCTCCGCCTGTTGGGGTTGGGTGCTTTGGGTTGGCAATTNGGGCAGGANTT 1479

>D105 draft CTTATAGGGTCTGGGATGGCCTCTCTTGGGACAGGGACNTTC-ACCTGCCAC 1533
    >D132-1 CTTATAGGGTCTGGGATGGCCTCTTTTGGGA▒▒GGGACCTTC-TCCTGCCAC 1529
     >132-2 [redacted] 1527
 >D64 draft CTTATAGGNTCTGGGATGSCYTCTTTTGGGACRGGGACYTYYNTCCTGYCAC 1528
      >D129 CTTATAGGGTCTGGGATGGCCTCTTTTGGGAC▒GGGACCTTC-TCCTGCCAC 1521
      >95-2 CTTATAGGGTCTGGGATGGCCTCTTTTGGGACGGGGACCTTC-TCCTGCCAC 1540
 >D95 (6D9) CTTATAGGGTCTGGGATGGCCTCTTTTGGGACGGGGACCTTC-TCCTGCCAC 674
      >95-1 CTTATAGGGTCTGGGATGGCCTCTTTTGGGNSTGGGACCTTC-TCCTGCCAC 1530

>D105 draft CGCACCGTTGGTTCTCCACGAGAGGCTTGGTGCAGACCAAGGATGCTGCGCT 1585
    >D132-1 CGCAC▒G▒TGGTTTTCCAC▒GGAGG▒TTGGTGCGGACCAAGGATGCTGCGCT 1581
     >132-2 [redacted]T▒TCCACG▒GAGGCTTGGTGC▒GACCAAGGATGCTGCGCT 1579
 >D64 draft CRCASCKCTGKTTTTCCATGRGARGCTTGGTGCGGACCAAGGATGCTGCGCT 1580
      >D129 CGCAC▒GCTGGTTTTCC▒C▒GGAGG▒TTGGTGCGGACCAAGGATGCTGCGCT 1573
      >95-2 CGCACCGCTGGTTTTCCATGGGAGGCTTGGTGCGGACCAAGGATGCTGCGCT 1592
 >D95 (6D9) CGCACCGCTGGTTTTCCATGGGAGGCTTGGTGCGGACCAAGGATGCTGCGCT 726
      >95-1 CGCACTGCTGGTTTTCCACAGGAGGTTTGGTGCGGACCAAGGATGCTGCNCN 1582

>D105 draft CTGAACCCCTATTAAACCTTGAACAATGCCTNTTNTGACCTTCCCAGGAGAG 1637
    >D132-1 CTGAACCCCTATTAAACCTTGAACAATGCCCCTT▒TGACCTTCCCA▒▒A▒AG 1633
     >132-2 ▒TGAACCCCTATTAAACCTTGAACAATGCC▒▒TT▒TGACCTTCCCAGGAGAG 1631
 >D64 draft CTGAACCCCTRTTAAACMTTGAACRCTGYCCCTTSYRACCTTCCCAGGAGAG 1632
      >D129 CTGAACCCCTATTAAACCTTGAACAATGCCCCTT▒TGACCTTCCCAG▒A▒AG 1625
      >95-2 CTGAACCCCTATTAAACCTTGAACGCTGTCCCTTGTGACCTTCCCAGGAGAG 1644
 >D95 (6D9) CTGAACCCCTATTAAACCTTGAACGCTGTCCCTTGTGACCTTCCCAGGAGAG 778
     >95-1 CTGAACCCCTATTAAACCTTGAACAATGCCCCTTCTGACCTTCCCANNANAN 1634
```

```
>D105 draft GAAACCCCAGGTTTTCTTCCCGTTTTTAGGGTAGTTTTCCTACTTTCTTCCC 1689
    >D132-1 GAAACCCCA■■TTTTCTTCCC■TTTTTAGGGTAGTTTTCCTACTTTCTT■■■ 1682
     >132-2 GAAACCCCAGGTTTTCTTCCCGTTTTTAGGGTAGTTTTCCTACTTTCTT■■■ 1683
 >D64 draft GAAACCCCAGGTTTTCTTCCCGTTSTTAGGGTAGTTTTCYTACTTTCTWNNN 1684
      >D129 GAAACCCCA■■TTTTCTTCCC■TTTTTAGGGTAGTTTTCCTACTTTCTT■■■ 1674
      >95-2 GAAACCCCAGGTTTTCTTCCCGTTCTTAGGGTAGTTTTCCTACTTTCTTCCC 1696
 >D95 (6D9) GAAACCCCAGGTTTTCTTCCCGTTCTTAGGGTAGTTTTCCTACTTTCTTCCC 830
      >95-1 GAAACCCCAACTTTTCTTCCCCTTTTTAGGGTANTTTTCCTACTTTCTT--- 1683

>D105 draft TTCTGCTCACTCCCGAGTGTGTCTGTGTTGCAGGTTTGAGGTGCGGGGAGGT 1741
    >D132-1 TTCTGCTCACTCCTG■GTGTGTCTGCATTGCAGGTTTGAGGTGCGGGGAGGT 1734
     >132-2 TTCTGCTCACTCCTGAGTGTGTCTG■■TTGCAGGTTTGAGGTGCGGGGAGGT 1735
 >D64 draft TTCTRCTCACTCCTGAGTGTGTNTGCATTGCAGGTTTGAGGTGCGGGGAGGT 1736
      >D129 TTCTGCTCACTCCTG■GTGTGTCTGCATTGCAGGTTTGAGGTGCGGGGAGGT 1726
      >95-2 TTCTACTCACTCCTGAGTGTGTCTGCATTGCAGGTTTGAGGTGCGGGGAGGT 1748
 >D95 (6D9) TTCTACTCACTCCTGAGTGTGTCTGCATTGCAGGTTTGAGGTGCGGGGAGGT 882
      >95-1 TTCTGCTCACTCCTGGGTGTGTCTGCATTGCAGGTTTGAGGTGCGGGGAGGT 1735

>D105 draft GGCCNTTGTGCGTACATAAACCNNCAGAAGATCAGCTCAGCCCTCTGCAACA 1793
    >D132-1 GGCCCTTGTGCGTACATAAACCAGGAGAAGATCAGCTCAGCC■TCTGCAACA 1786
     >132-2 ■GCCCTTGTGCGTACATAAACCAG■AGAAGATCAGCTCAGCC■TCTGCAACA 1787
 >D64 draft GGCCCTTGTGCRTACATMMMCCAGGAGANNNTCAGCTCAGCCYTNTGNAAMA 1788
      >D129 GGCCCTTGTGCGTACATAAACCAGGAGAAGATCAGCTCAGCC■TCTGCAACA 1778
      >95-2 GGCCCTTGTGCGTACATAAAC---CAGGAGATCAGCTCAGCCTTCTGCAACA 1797
 >D95 (6D9) GGCCCTTGTGCGTACATAAACCAGGAGA---TCAGCTCAGCCTTCTGCAACA 931
      >95-1 GGCCCTTGTGCGTACATAAACCAGGAGAAGRTCAGCTCAGCCTTCTGCAACA 1787

>D105 draft CGGAGAAATACTGGATC                                   1810
    >D132-1 CGGAGAAATACTGGATC                                   1803
     >132-2 CGGAGAAATACTGGATC                                   1804
 >D64 draft CGGAGAAATACTGGATC                                   1805
      >D129 CGGAGAAATACTGGATC                                   1795
      >95-2 CGGAGAAATACTGGATC                                   1814
 >D95 (6D9) CGGAGAAATACTGGATC                                   948
      >95-1 CGGAGAAATACTGGATC                                   1804
```

Appendix 4: Alignment of genomic sequences, including draft sequences from some ducks. Ns represent unknown sequence. Shading indicates mismatches. Dashes indiates gaps introduced to maximize the alignment.