Genomic Sequence and Evolutionary Analysis of the 3' End
of the Murine T Cell Receptor γ Locus

by

Michael John Parlee
Bachelor of Science, University of Victoria, Victoria, B.C., Canada, 1993

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Biology

We accept this dissertation as conforming
to the required standard

[redacted]

Dr. B. F. Koop, Supervisor (Department of Biology)

[redacted]

Dr. N. Sherwood, Departmental Member (Department of Biology)

[redacted]

Dr. C. Upton, Outside Member (Department of Biochemistry and Microbiology)

[redacted]

Dr. L. Mizzen, External Examiner (StressGen, Inc.)

© Michael John Parlee, 1997

University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by
photocopying or other means, without the permission of the author.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I thank my supervisor, Dr. Ben Koop, for providing the opportunity for me to do this work and for introducing me to molecular evolution and computational biology. I thank Linda McKinnel and Ute Rink for introducing me to molecular techniques. Lorelei Lew and Jacquie Brinkman I thank for their tremendous help in obtaining much of the sequence data reported here. I thank all of my co-workers at the Centre for Environmental Health for all of their advice, camaraderie and useful discussions. I also thank my wife, Larissa. Her patience, support, and editorial skills were indispensable in writing this thesis. Most importantly, I thank my parents.

INTRODUCTION

The immune system is a critical component of all vertebrate organisms. In distinguishing self from non-self, it plays a important role in maintaining the integrity of components and systems that make up the living organism. The immune system is an extremely complex system consisting of many levels of both effectors and regulatory pathways. These levels range from a broad spectrum innate response to a highly specific acquired response, in which a remarkable system for generating diversity in the immune response enables a specific response to be mounted against any infectious agent that might be encountered. With such extensive defenses, it is important that the organism prevent its immune system from attacking other self components. The body does this by educating the immune system, destroying effector cells that react with self components as they develop. This system works very well for the most part but occasionally the mechanisms of self tolerance are bypassed and the immune system begins to destroy the body. In humans this causes several debilitating conditions known as autoimmune diseases such as systemic lupus erythematosus, multiple sclerosis and arthritis. The degenerative nature of these diseases makes them costly both in monetary terms and in human suffering. A better understanding of immune system function, organization and regulation will improve our understanding of the causes of these diseases and allow the development of cures or treatments to lessen their impact on human health (Austyn and Wood, 1993; Roitt, 1991).

At the first level of host defence are the physical and chemical barriers. The dry, acidic condition of the skin, for example, presents an inhospitable climate for most microorganisms. Another barrier is the constant washing of the eyes by tears, which sweep out foreign particles and contain lysozyme to break down the cell walls of bacteria. The mouth and sinuses are protected by mucous that traps microorganisms and allows them to be expelled by sneezing or coughing. Cilia in the trachea constantly sweep mucous and trapped foreign particles away from the lungs. Additionally, the normal body flora serves as a protective barrier to infection by competing for nutrients and by producing chemical

secretions that inhibit the growth of competing microbes. Injuries that penetrate the skin, surgery, burn damage and antibiotic treatment that kills the normal flora, all greatly increase the chance of infection by organisms commonly present in the environment. These infections are termed opportunistic infections because the organisms involved are not specialized to bypass the physical defences, but cause infection only when given the opportunity. Other organisms depend on infecting a host organism as part of their life cycle and are specialized to bypass physical and chemical barriers. These organisms commonly infect via the mucosal layers of the mouth, nose, throat, bronchi, and urogenital tracts. Vector borne pathogens are aided in bypassing physical and chemical barriers by intermediate host organisms in which a portion of their life cycle is spent (Austyn and Wood, 1993; Roitt, 1991).

When a microbe bypasses the physical barriers to infection, the body is ready with a more directed defence response. The immune system employs both an innate response and an acquired response. The innate response is of broad specificity and recognizes common properties of invading microbes, whereas the acquired response is highly specific and responds to unique sites on foreign molecules. In order to recognize specific sites on any given invader, the antigen receptors of the acquired immune response require a diverse array of binding specificities.

The innate response involves phagocytic cells and a series of proteins known as complement. Complement is made up of about 20 plasma proteins and is activated by a cascade in which each activated component, in turn, activates another. This cascade results in the coating of the infectious agent by complement proteins and may result in the lysis of susceptible bacteria through the formation of a membrane attack complex. If it is unable to lyse the target microbe, the coating of complement still serves to facilitate the adherence of the microbe to phagocytic cells. The cleaved byproducts of the complement cascade also aid in the phagocytic response both as chemoattractants and by signalling the phagocytic cells

to upregulate their oxidative metabolism and complement receptors. Complement is the first line of defence against microbes that have infiltrated the physical defences of the body. It is ubiquitous in the plasma and responds immediately to infection, tagging the invading particle for removal by phagocytes (Austyn and Wood, 1993; Roitt, 1991).

There are two types of phagocytic cells in the vertebrate immune system. One type of phagocyte, the polymorphonuclear neutrophil (PMN), is named for its amorphous, multilobed nucleus. The other major type of phagocytic cell is the monocyte. Monocytes differentiate into macrophages or other cell types depending on the tissue in which they reside. In order for phagocytosis to occur, the invading microbe must adhere to the phagocytic cell. The cell membrane of the phagocyte then extends around the microbe engulfing it completely, ingesting the microbe and subsequently destroying it with a battery of oxidative agents, low pH and lytic enzymes. The innate immune response is non-specific, depending only on the capacity of a microbe to activate a complement cascade or to bind the cell membrane of a phagocyte (Austyn and Wood, 1993; Roitt, 1991).

The innate response alone is not sufficient to protect the body from all infectious agents. Mice with severe combined immunodeficiency (SCID) have a functional innate immune response, but lack an acquired immune response. SCID mice die within days of their birth if they are not reared under sterile conditions. The acquired immune response can be broken down into two arms. Humoral immunity is mediated by B lymphocytes, and involves a direct response to foreign molecules or exogenous antigen (Ag). Cellular immunity is effected by T lymphocytes, which typically respond to antigen that has first been processed and displayed on the surface of host cells with major histocompatibility complex (MHC) proteins. A close interaction occurs between both arms of the acquired immune system and between the acquired and innate responses as well (Austyn and Wood, 1993; Roitt, 1991).

The humoral immune system surveys the blood and interstitial spaces for infection and responds with the production of antibodies tailored to bind specifically to a given antigen. The effector cells of this response are known as B lymphocytes or B cells. B cells are stimulated when their surface receptors, called B cell receptors, bind to a corresponding target, called an antigen. Each B cell expresses a B cell receptor specific for just one site on an antigen. This site is called an epitope. During B cell development, a B cell receptor is produced through a genetic recombination event. Stimulation of the B cell by antigen binding results in the secretion of a soluble form of the B cell receptor. These multivalent antigen binding proteins are called antibodies or immunoglobulins (Igs). Additionally, the stimulated B cell undergoes a series of cell divisions resulting in a clonal expansion, greatly increasing the number of B cells with specific binding affinity to the original antigen. Coating of the antigen with antibody then facilitates its removal by phagocytic cells and complement mediated pathways (Austyn and Wood, 1993; Roitt, 1991).

The cellular immune response is mediated by T cells, which recognize antigen through a surface receptor called a T cell receptor (TCR). The TCR is evolutionarily related to the Ig molecules of the B cells and displays some similarity in structure as well as some differences. The TCR is a heterodimeric protein, meaning that a single receptor is composed of two different protein subunits. In humans and mice, most T cells express a TCR composed of one α subunit and one β subunit. These cells are called $\alpha\beta$ T cells. A smaller population of T cells express a TCR composed of one γ subunit and one δ subunit, and these cells are known as $\gamma\delta$ T cells. While the function of the $\alpha\beta$ T cells is well studied, the function of the $\gamma\delta$ T cells is not yet understood. Both types of T cells share common characteristics and it was assumed that both the $\alpha\beta$ and $\gamma\delta$ T cells would share similar functions. However, evidence now suggests that the function of the $\gamma\delta$ T cells is distinct from that of the $\alpha\beta$ T cells (Kaufmann, 1996).

Unlike B cells, which recognize exogenous antigen, T cell receptors typically bind processed antigen presented by MHC proteins on the surface of host cells. T cells respond to antigen in two distinct ways. T helper cells (Th) are distinguished by the presence of CD4 surface markers (CD4+). The Th cells recognize antigen presented by MHC class II molecules and respond by secreting various growth and differentiation factors to aid other cells of the immune system in battling the infection. The other subset expresses the surface marker CD8 (CD8+), and are called cytotoxic T cells (Tc). They recognize antigen expressed with MHC class I and respond by directly killing the infected host cell to prevent the spread of the infection. The requirement of a particular MHC class for a particular response is known as MHC restriction (Ausyn and Wood, 1993; Roitt, 1991).

The phenomenon of MHC restriction is well studied in the αβ T cells, but a growing body of evidence suggests that most γδ T cells are not restricted in the same manner. The objective of the work presented here is to provide new clues about the evolution and function of the γδ T cells and how it differs from our current understanding of αβ T cell structure and function.

1.1 Immunoglobulin Structure

The TCRs and Igs belong to a large family of related proteins known as the Ig superfamily. The Ig superfamily is most clearly characterized by a feature of the secondary structure known as an Ig fold. The protein domain in which this feature occurs in known as an Ig domain. The Ig fold is usually stabilized by an intrachain disulphide bond between conserved cysteine residues. The presence of the regularly spaced cysteine residues is not a requirement for the formation of an Ig fold, but it is a common feature that has been used in the classification of Ig superfamily members. Amino acid sequence homology, secondary structure, location, and function of a protein are all considered for the inclusion of that protein in the Ig superfamily. In addition to the Ig, TCR and several other lymphocyte surface antigens, this diverse family of proteins includes neural cell adhesion

molecule (N-CAM), myelin associated glycoprotein (MAG), platelet derived growth factor receptor (PDGFR), macrophage Fc receptor (FcR) and carcino-embryonic antigen (CEA). Most of the Ig superfamily members are expressed on the surface of cells and function in cellular interaction through specific and often homophillic ligand binding (Williams, 1987). The TCR and Ig have branched from the Ig superfamily to augment the vertebrate immune system with the ability to recognize a vast number of ligands. It is due to their common heritage that the TCR and Ig share some basic structural features and very similar gene organization.

Ig domains can be classified as either variable (V) or constant (C) domains. The basic features of the Ig fold are shared by both the V and C domains. Two β-sheets with alternating hydrophobic and hydrophillic residues, the hydrophobic residues facing inward, are stabilized by a conserved disulphide bond. The V domain differs from the C domain in that the V domain contains an extra loop in the middle of the fold which, in antibodies, contains a hypervariable region. TCRs and Igs both have V domains at the amino termini of all their subunits whereas C domains occur toward the carboxyl termini. In both molecules it is the V domain that is responsible for ligand binding whereas the more conserved C domain is involved in structure and signal transduction. The C domain of antibodies is recognized by the macrophage FcR to enhance phagocytosis of foreign microbes (Williams, 1987).

The structures of the TCR and Ig molecules share similar motifs. Igs are made up of two identical light chains (Ig L) and two identical heavy chains (Ig H) linked together by interchain disulphide bonds (Figure 1.1). The light chains consist of one V and one C domain, whereas the larger, heavy chains have one V and three C domains. TCRs, on the other hand, consist of two different subunits, either α and β or γ and δ, linked by interchain disulphide bonds. Each subunit contains a single V domain and a single C domain. Each

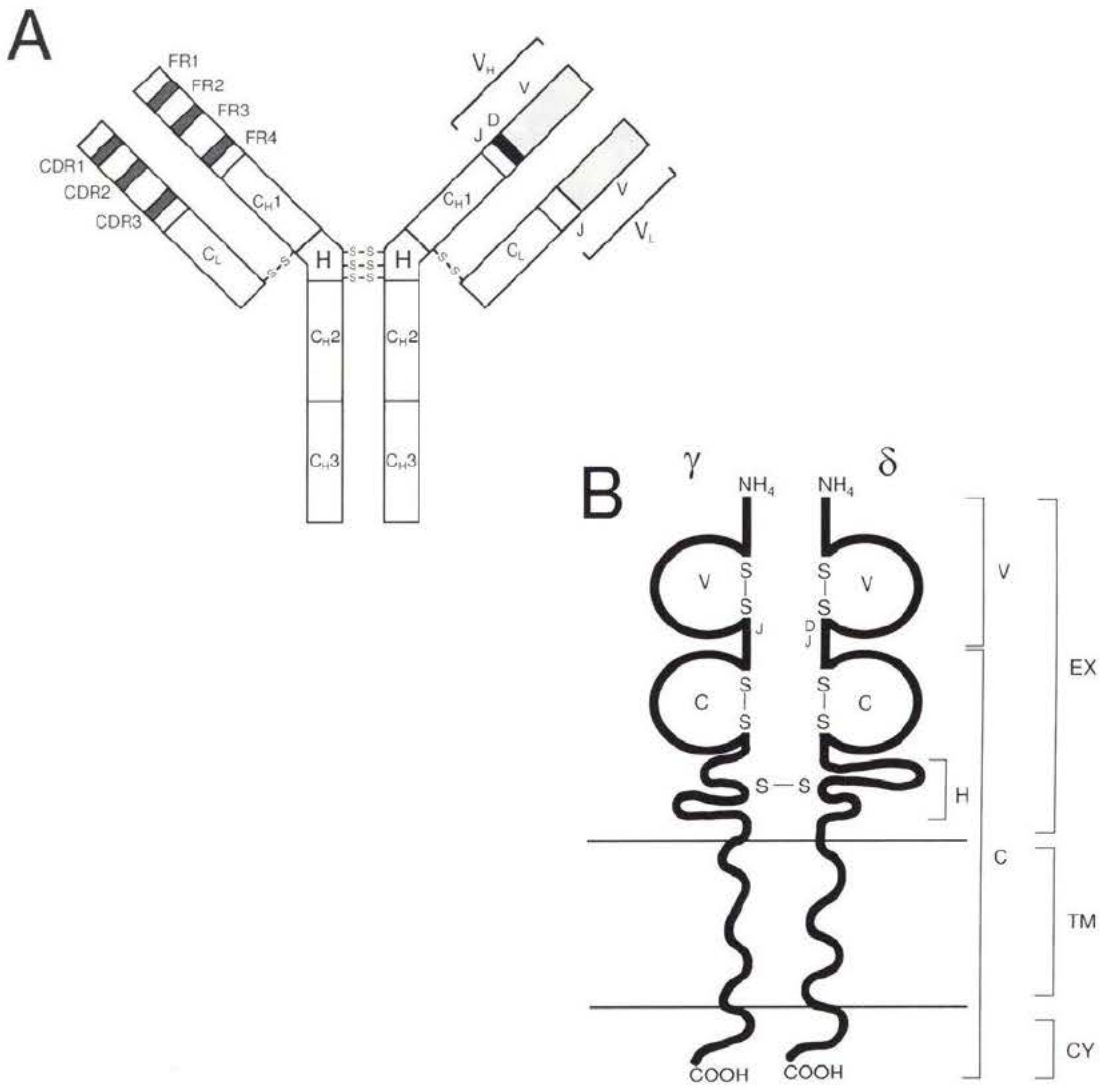

Figure 1.1. A prototypical immunoglobulin molecule and γδ T cell receptor. A. Immunoglobulin heavy (H) chains are held together by three disulphide bonds. Shorter light (L) chains are associated with the H chains by single disulphide bonds. Variable (VH and VL) and constant regions (CH and CL) regions of the H and L chains are indicated. Variable (V), diversity (D) and joining (J) gene segments encode the regions within the V region marked V, D and J respectively. Complementarity determining regions (CDR) are indicated by dark bands between the framework regions (FR). A hinge region (H) falls between the constant (C) domains of the heavy chains. Modified from Lieber (1991). B. γδ T cell receptor immunoglobulin folds are indicated by loops stabilized by disulphide bonds. The variable domain, encoded by V, D and J gene segments is denoted by the letter V. The constant region is denoted by the letter C, and the hinge region within the C domian is denoted by the letter H. EX indicates the extracellular region of the receptor, TM indicates the transmembrane region, and CY indicates the cytoplasmic region. Modified from Williams (1987).

TCR is monovalent in binding capacity, whereas the basic structural unit of the Igs includes two binding sites per molecule. The multivalency of the Igs is further compounded in various classes of secreted antibodies, in which several molecules of the prototypical IgG antibody described above are linked together at their constant regions (Tonegawa, 1983; Lieber, 1991; Roitt, 1991).

At the level of the gene, the TCR and Ig loci share a very similar organization. The V domains of the TCR β and δ subunits and the Ig H chains are encoded by variable (V), diversity (D) and joining (J) gene fragments, whereas the V domains of the TCR α and γ subunits and the Ig L chains are encoded by just a V and a J gene segment (Figure 1.1). The V gene segment encodes the majority of the V region and is joined to the constant region by a small J gene segment. The D segment encodes a small region between the V and J segments in those subunits in which it is present. The V domains of the antigen receptors contain three regions of high diversity. These regions are termed complimentary determining regions (CDR) while the surrounding regions of the V domain are termed framework regions (FR). The CDRs line the binding pocket of the receptors and contribute to the diversity of antigen which can be recognized. CDR3 corresponds to the V-J or V-D-J gene segment junctions and is the most variable of the three CDRs. Within the C domains of the TCRs and Igs is a short hinge domain within which disulphide bonds link the two Ig H chains in Igs and the two subunits of the TCRs. In the TCRs, the hinge domain occurs between the extracellular and transmembrane domains of the C region. In the Igs the hinge domain occurs at approximately 1/3 of the length of the heavy chain C region at the base of the two antigen binding domains. TCRs and Igs share common structural motifs as well as common functional motifs. This is further exemplified in a shared system for generating diversity (Tonegawa, 1983; Davis and Bjorkman, 1988; Lieber, 1991).

1.2 Generation of Diversity in the TCR and Ig Antigen Receptors

In order to protect the body from infection, the immune system must be capable of recognizing and responding to a large number of infectious agents. In addition to existing pathogens, it would be advantageous for an organism to be able to respond to newly occurring or mutated pathogens as well. In order to provide this protection without necessitating a single gene for each unique receptor, the genes encoding the antigen receptors in vertebrates have evolved a mechanism for generating diversity through somatic rearrangement, N region diversity and hypermutation. This mechanism enables the immune system to mount a specific attack against virtually any invader, whether the site of infection be humoral or intracellular. Rather than individually encoding a receptor of each possible specificity, all of the specificities required of a TCR or Ig subunit are encoded at a single genomic locus. Diversity is primarily generated by utilizing different combinations of V, D and J gene segments, which can number anywhere from one to several hundred depending on the locus and type of gene segment (Tonegawa, 1983; Davis and Bjorkman, 1988).

In the prototypical organization of an antigen receptor locus of a higher vertebrate, one finds a number of V gene segments at the 5' end of the locus. The string of V segments is then followed by a lesser number of D gene segments, if present, and a number of J gene segments. The J gene segments are then followed by a C gene, which may be followed by more D and J segments and a second C gene. The gene segments are so named because, unlike true exons, they do not possess both donor and acceptor splice sites for RNA processing. In lymphocytes developing in the bone marrow and in the thymus, a single set of gene segments is brought together from the total repertoire by DNA recombination events to form a single gene (Figure 1.2). Each V gene segment possesses a promoter and a transcription start site at its 5' end. At the 3' end of each V gene segment, as well as at the 5' end of each D and J segment, lies a recombination signal sequence. At the 3' end of each D segment is another recombination signal which is matched by another

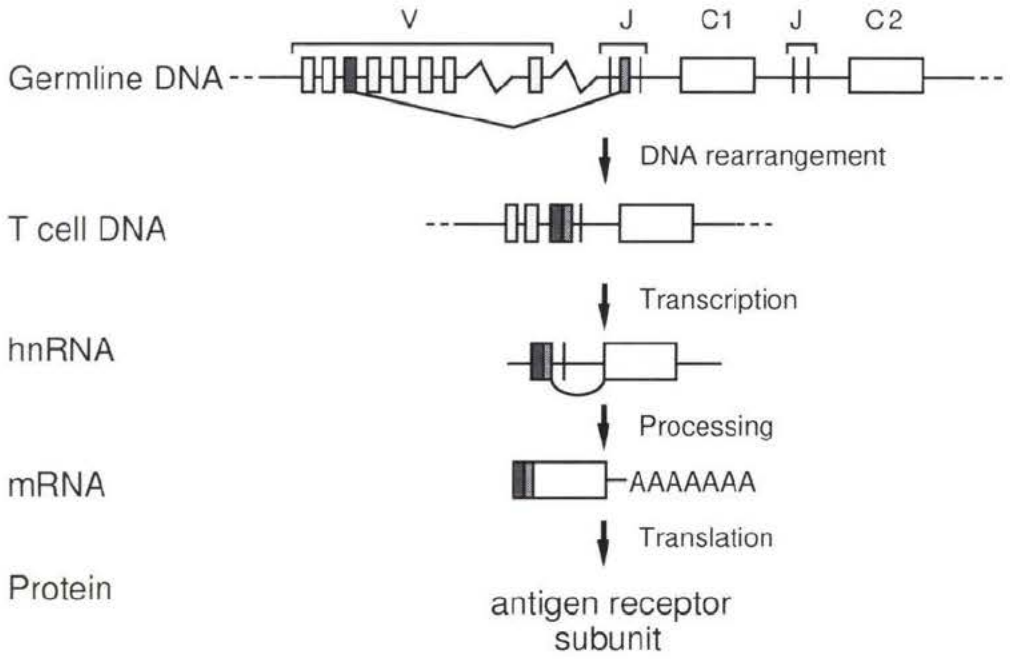

Figure 1.2. V(D)J recombination. Signal sequences 3' of V gene segments and 5' of J gene segments mediate a recombination event in the germline DNA in which a single V gene segment and a single J gene segment are brought together and the intervening DNA is deleted. The joined V and J gene segments form a single exon encoding the variable domain of the antigen receptor. This exon is transcribed together with the exons of the neighbouring constant (C) region. Intron removal and exon splicing produce mRNA that is then translated to form an antigen receptor subunit.

signal sequence at the 5' end of each J segment. At the 3' end of each J gene segment, however, lies a normal donor splice site which is matched by an acceptor splice site at the 5' end of the first exon of the C gene. An enzyme with recombinase activity matches the recombination signal sequences of one V and one D or J gene segment from the total repertoire and removes the intervening DNA, leaving the V segment joined directly to a D or J gene segment. If a D gene segment were used, a second recombinase step between the D and a J gene segment is necessary to form a complete gene. Once the rearrangement is complete the antigen receptor gene is transcribed to RNA, which is processed to remove introns, and the resulting mRNA is translated to produce a receptor of a unique specificity. This mechanism is known as V(D)J recombination and the diversity due to the combining of different gene segments is referred to as combinatorial diversity (Tonegawa, 1983).

The mechanism through which recombination occurs is a specific enzyme mediated event. The recombination signal sites bordering each gene segment are highly conserved. V and D gene segments are followed by a signal sequence of a conserved heptamer, CACAGTG, and a nonamer, ACAAAAACC, separated by a spacer sequence of either 12 or 23 base pairs. D and J gene segments are immediately preceded by a nonamer, GGTTTTTGT, and a heptamer, CACTGTG, also separated by either 12 or 23 base pairs of spacer sequence (Tonegawa, 1983). During the recombination process cuts are made at the ends of each signal directly adjacent to the heptamer. The two signal ends are then joined together forming a signal joint and the ends of the two gene segments are joined to form a coding joint. Prior to the formation of the coding joint, the diversity of the antigen receptors is further increased due to the random deletion of nucleotides from the end of each gene segment. Nucleotide deletion is followed by template independent addition of nucleotides at the ends of the gene segments by terminal deoxynucleotidyl transferase (TdT) which further increases the diversity of the products (Lieber, 1991). The

contribution of random deletion and addition to the antigen receptors during VJ joining is termed N region diversity (Roitt, 1991).

A third method for generating diversity, somatic hypermutation, has been observed in the complementarity determining regions (CDR) of the V gene segment in immunoglobulins, but this has not been observed in the TCR genes. The diversity generated by these mechanisms is further compounded by the necessity of joining two such subunits to form a single antigen binding receptor (Tonegawa, 1983; Davis and Bjorkman, 1988).

The effect of these combined mechanisms in generating diversity in the antigen receptors can be estimated arithmetically. The mouse TCR α locus, for example, contains at least 70 V gene segments and 61 J gene segments (Table 1.1). Multiplying these together gives a total of 4270 different combinations. Random deletion of nucleotides from the ends of the gene segments usually removes seven or fewer base pairs (Lieber, 1991). Thus, for each recombination event the number of unique TCRs can be multiplied by seven for the end of each of the gene segments involved. TCR α has only V and J gene segments so the diversity contributed by deletion would be 49. Random addition of nucleotides to the ends of the gene segments increases diversity by a factor of four for each position up to the maximum length of the addition. This can be calculated by summing 4^n from 0 to n, where n is the maximum number of nucleotides added. Assuming a maximum of six nucleotides is added to the TCR, then 5461 unique combinations are added. These numbers do not take into account those products that are nonfunctional due to a shift in the reading frame. Assuming that the number of nucleotides added and deleted is truly random and there is only one open reading frame, 1/3 of the products would be functional. Codon redundancy must also be considered. Only 1/3 of the possible nucleotide deletions and additions would yield unique translation products because 20 amino acids are encoded by 61 codons. Even with this ninefold reduction, the potential diversity of the TCR α subunit

Table 1.1. Potential sequence diversity of the T cell receptor genes of adult mice. From Lieber, 1991.

	α	β	γ	δ
V	70	25	7	8
D	-	2	-	2
J	61	12	4	2
Nt addition	5461	5461^{2}	5461	5461^{4}
Nt loss	7^{2}	7^{4}	7^{2}	7^{6}
D-frame	-	1	-	1
J-frame	1/3	1/3	1/3	1/3
Codon Redundancy	1/3	1/3	1/3	1/3
Combinations of chains	1.0×10^{8}	4.8×10^{12}	8.3×10^{5}	3.7×10^{20}
Total combinations of Ag receptors	5.8×10^{20}		3.1×10^{26}	

alone is 1.0×10^8. When this is multiplied by the diversity of the β subunits (4.8×10^{12}), the possible number of unique αβ TCR receptors is on the order of 10^{20} (Lieber, 1991). This is particularly impressive when one considers that there are probably only 50 to 100 thousand genes in the entire human genome. Through this system of combinatorial and junctional diversity, the entire repertoire of receptors that a vertebrate will need to fight infection throughout its life is provided with a comparatively minuscule amount of coding material.

In generating many millions of unique antigen receptors, it is impossible to avoid receptors that recognize components of the host's own body. To prevent an autoimmune response, the body removes lymphocytes that react with self components as soon as they develop. In the thymus, newly developed T cells are tested for self reactivity by macrophages which present self antigen on their cell surface with MHC class II. Those T cells that recognize self molecules are killed by a mechanism of programmed cell death called apoptosis. B cells undergo a similar selection process in the bone marrow. Some lymphocytes may be only weakly reactive to self components and thus escape the selection process. A later encounter with an exogenous antigen that induces a strong response may sensitize these previously innocuous lymphocytes causing them to respond more strongly to the self antigen. The result is an immune response against components of the organism's own body.

1.3 γδ T Cell Receptors

According to the classic model of TCR function, αβ TCRs recognize antigen presented to TCRs in the peptide binding cleft of MHC proteins on the surface of host cells. The antigen presented to αβ TCRs is termed processed antigen because the original antigen is cleaved into short peptides by the host cell. These peptides are introduced into vesicles containing newly formed MHC molecules. The processed peptides associate with

a peptide binding cleft in the MHC molecule and the MHC-antigen complexes proceed to the cell membrane where they can interact with TCRs. There are two classes of MHC. MHC class I interacts with $CD8^+$, or cytotoxic, T cells whereas MHC class II interacts with $CD4^+$, or helper, T cells. MHC class I is found on most of the cells in the body. It presents antigen derived from self components and intracellular pathogens such as viruses and mycoplasmas. When a Tc cell recognizes an antigen presented on an MHC class I molecule, it responds by inducing the death of the infected host cell. MHC class II molecules are displayed primarily on specialized antigen presenting cells, such as macrophages, PMNs and B lymphocytes. As they phagocytize and kill invaders they process antigens from the invading particle and display the processed peptides with MHC class II on their cell surface. When a Th cell recognizes antigen presented on a class II molecule it responds with the production of lymphokines in order to coordinate an immune response to the site of infection (Germain, 1994).

In human and mouse more than 90% of T lymphocytes express the αβ TCR. Less than 10% of T lymphocytes in these organisms express the γδ TCR. The αβ T cells have been extensively studied and it is from this subset that our understanding of T cell function comes. The smaller subset of lymphocytes expressing the γδ TCR shares many characteristics with the αβ T cells, but an ever increasing body of work has shown that the γδ T cells do not adhere to the established paradigm of T cell function (Kaufmann, 1996; Haas *et al.*, 1993).

Discovery and initial work on the γδ TCR came primarily out of the laboratories of Tonegawa (Heilig and Tonegawa, 1986) and Raulet (Garman *et al.*, 1986). Each of these groups proposed its own nomenclature for the TCRγ gene segments, and yet another group proposed a third (Traunecker *et al.*, 1986). To settle the nomenclature problem, the WHO-IUIS subcommittee on TCR nomenclature has proposed a standard naming convention but

this has only been applied to the Vγ gene segments so far. The nomenclature of Garman *et al.* (1986) will be used throughout this thesis (Table 1.2).

Lymphocytes expressing γδ TCRs are classified as T cells due to the homology of the γδ and αβ antigen receptors, the expression of common surface antigens and development in the thymus. The more that is learned about these T cells, however, the more distinct they appear to be from T cells expressing αβ TCRs. In the mouse, γδ T cells appear earlier in fetal development than do αβ T cells. They arise in waves of distinct clonalities and they may not require MHC restriction. Additionally, γδ T cells expressing certain Vγ gene segments have been found to preferentially localize to particular tissues in the mouse. These differences suggest a function distinct from that of the αβ T cells, but as yet the function of the γδ T cells remains unknown.

1.4 Ontogeny of the γδ T cells

The γδ T cells appear earlier in fetal ontogeny than do αβ T cells. In the mouse fetus, TCRγ mRNA is detectable in liver and gut tissue by day 11 of fetal development. TCRγ mRNA is first detected in the thymus at day 13. In comparison, TCRβ mRNA first appears in the thymus at day 15 and TCRα at day 16 (Carding *et al.*, 1990; Haars *et al.*, 1986). In conjunction with the early appearance of γδ T cells in the developing fetus, γδ T cell clones utilizing different Vγ gene segments have been shown to appear in the thymus in distinct waves (Figure 1.3). Thymocytes expressing Vγ3 gene segments appear before day 14 and remain dominant until day 17 when their numbers drop sharply to disappear by day 18. Starting at day 14 γδ cells expressing Vγ4 gene segments begin to appear and rise rapidly in number until day 18 when they begin to drop to their lower postnatal level. At day 15 cells expressing Vγ2 gene segments first appear, rise to a peak by day 19, then drop slightly in number but remain the dominant γδ cell population in the postnatal and adult thymus (Havran & Allison, 1988; Ito *et al.*, 1989). The underlying reason for this

Table 1.2. Comparison of the three most common nomenclatures for the V gene segments of the mouse TCR γ locus. From Arden *et al.*, 1995.

WHO-IUIS	Heilig and Tonegawa (1986)	Garman et al. (1986)
GV1S1	Vγ5	Vγ3
GV2S1	Vγ6	Vγ4
GV3S1	Vγ4	Vγ2
GV4S1	Vγ7	Vγ5
GV5S1	Vγ1	Vγ1.1
GV5S2	Vγ2	Vγ1.2
GV5S3	Vγ3	Vγ1.3

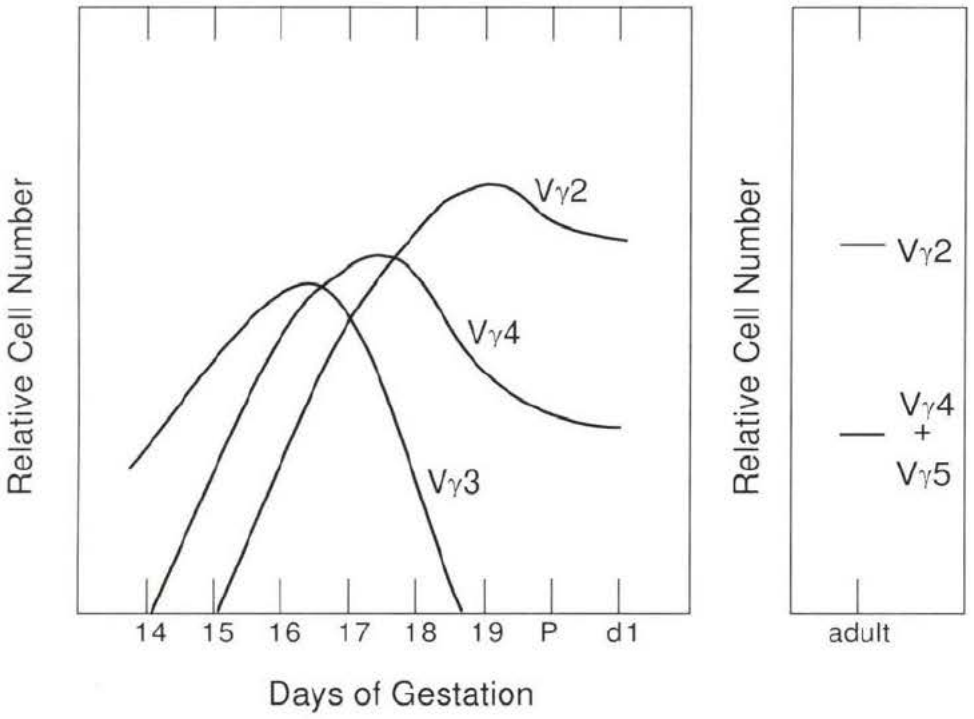

Figure 1.3. Relative levels of Vγ clonalities in thymus during fetal development. Thymocytes expressing specific Vγ gene rearrangements arise in waves in the fetal thymus. The relative cell numbers in each clonal population in the thymus from day 14 of fetal development to the first day (d1) after parturition (P) are shown. Relative numbers of specific Vγ clones in the adult thymus are also shown. Adapted from Vernooij (1992).

temporal pattern of expression is not known. Interestingly it reflects the order of the V gene segments in the genome (Vernooij, 1992).

The localization of T cell clones expressing a particular TCR γ rearrangement to a particular tissue type occurs in mice but has not been observed in humans. In the mouse, there are three distinct subsets of the γδ T cell, the first of which is characterized by tissue localization. The first subset consists of the γδ cells that appear in the fetal thymus and express Vγ3 and Vγ4 gene segments (Havran & Allison, 1988; Ito *et al.*, 1989; Itohara *et al.*, 1989; Lafaille *et al.*, 1989). In the mature adult, Vγ3 T cells are localized in the skin and T cells expressing Vγ4 are found in the mucosal surfaces of the uterus, vagina and tongue (Asarnow *et al.*, 1988; Itohara *et al.*, 1990). The VJ junctions of this subset are devoid of junctional diversity and the TCRs of these cells are thought to be monospecific in their ability to bind antigen. The second subset are those T cells populating the thymus of the adult mouse. γδ T cells expressing Vγ2 and Vγ1.1 gene segments predominate in this subset while a few express Vγ1.2 and Vγ5 gene segments (Itohara *et al.*, 1989; Takegaki *et al.*, 1989a; Bluestone *et al.*, 1991). Unlike the TCRs of the γδ T cells originating in the fetal thymus, the TCRs of the γδ cells in the adult thymus show an extremely high level of junctional diversity (Elliot *et al.*, 1988; Takegaki *et al.*, 1989a; Ezquerra *et al.*, 1990). Like αβ T cells, adult thymic γδ T cells circulate in the peripheral blood and are present in classic lymphoid organs such as the spleen and lymph nodes (Bluestone *et al.*, 1991). The third subset of γδ T cells matures independently of the thymus (Mosley *et al.*, 1990; Lefrancois *et al.*, 1990; Bandeira *et al.* 1991; Guy-Grand *et al.*, 1991). These cells express Vγ5 and Vγ1.1 gene segments. They exhibit high junctional diversity and are localized primarily in the intestinal epithelium (Bonneville *et al.*, 1988; Takagaki *et al.*, 1989b; Kyes *et al.*, 1989; Haas *et al.*, 1993; Kaufmann, 1996).

1.5 Antigen Recognition by γδ T Cells

One of the great advances in immunology was the elucidation of the mechanism of MHC restriction and the binding complex of TCR, antigen, and the MHC presentation molecule (Germain, 1994). However, it has been found that not all T cells are restricted in this manner. Most γδ T cells (Kozbor *et al.*, 1989; Holoshitz *et al.*, 1989; Matis *et al.*, 1989; Guo *et al.*, 1995) and a small population of αβ T cells do not exhibit MHC restriction. These T cells lack both CD4 and CD8 surface molecules and are therefore referred to as double negative (DN). DN T cells do not require either MHC class I or MHC class II for antigen binding. Instead, they have been found to recognize alternative antigen presenting molecules, such as CD1, H-2M3, and Qa-1, which lack the polymorphism of the MHC. These molecules show similarity to MHC class I and are expressed with β_2-microglobulin (β_2m) as are MHC class I molecules. They are encoded outside of the MHC locus and, due to their limited repertoire, can present only a few specific ligands (Porcelli, 1995; Beckman and Brenner, 1995; Kaufmann, 1996). Although there is evidence that γδ T cells recognize the heat shock protein (Hsp) 60 presented by the MHC class I-like protein, Qa-1, virtually normal levels of γδ T cells have been observed in β_2m deficient mice (Vidovic *et al.*, 1989; Imani and Soloski, 1991; Soloski *et al.*, 1995). Since these mice lack MHC class I or like molecules, they are not able to positively select T cells which recognize MHC and show greatly depressed levels of αβ T cells. That normal levels of γδ T cells were observed in these mice suggests that most γδ T cells may recognize antigen independently of MHC class I-like molecules all together. It is also possible that there may be an undiscovered non-MHC antigen presenting molecule involved in the selection of most γδ T cells, though as yet there is no evidence for such a molecule. Alternatively, γδ T cells may not require positive selection as do the αβ T cells, and thus are not affected by the lack of MHC (Raulet, 1994). There are also several lines of evidence that support antigen recognition by γδ T cells completely out of the context of antigen presentation molecules (Schild *et al.*, 1994; Weintraub, 1994; Kaliyaperumal, 1995).

The discovery that some human γδ T cells could recognize non-proteinaceous ligands (Pfeffer *et al.*, 1990) provided further evidence that there were exceptions to classic theory of antigen recognition by TCRs. Non-proteinaceous ligands have now been isolated from *Mycobacterium tuberculosis* as well from as several other bacteria and protozoa (Modlin *et al.*, 1989; Kabelitz *et al.*, 1990; Munk *et al.*, 1990; De Libero *et al.*, 1991; Behr and Dubois, 1992). The essential component of these non-peptide ligands is a phosphate bound to an alkyl, carbohydrate or nucleotide residue (Schoel *et al.*, 1994; Tanaka *et al.*, 1994, 1995; Constant *et al.*, 1994; Bürk *et al.*, 1995). All that is known about the presentation of these antigens is that they must be bound to the host membrane (Morita *et al.*, 1995). Recently a small subset of αβ T cells were also found to bind non-peptide antigens (Porcelli *et al.*, 1992; Beckman *et al.*, 1994; Sieling *et al.*, 1995). This subset of $CD4^-/CD8^-$ T cells recognized lipoarabinomannan and mycolic acids from mycobacterial cell walls that were expressed on host cell surfaces on MHC Class I-like CD1 molecules (Porcelli *et al.*, 1992; Beckman *et al.*, 1994; Sieling *et al.*, 1995). The activation of certain subsets of T cells by non-proteinaceous ligands provides yet another example of unusual activity that does not utilize the classic MHC pathway.

Some insight into the function of γδ T cells was gained when it was discovered that mycobacterial Hsps (heat shock proteins) elicited a strong T cell response (O'Brien *et al.*, 1989; O'Brien and Born, 1991; O'Brien *et al.*, 1991). In the mouse, Hsp-reactive γδ T cells typically express Vγ1.1 with high junctional diversity (O'Brien *et al.*, 1992). Hsps are so named because they were initially discovered when bacteria were exposed to heat shock. They are found in all prokaryotic and eukaryotic cells and they are among the most highly conserved of all proteins. The close similarity between Hsps led researchers to ask whether there could be a T cell response mounted against autologous Hsps in addition to those of bacterial origin (Born *et al.*, 1990). This is supported by the finding that γδ T cells in the skin of the mouse, known as dendritic epithelial cells (DECs), respond to a molecule

with little or no polymorphism expressed on keratinocytes following heat shock (Havran *et al.*, 1991; Allison and Havran, 1991). Kaufmann (1996) suggested that autologous Hsp produced in response to a variety of cell stresses may stimulate a broad response from γδ T cells. The fact that a diverse T cell response occurs specifically to a unique core sequence of Hsp60 suggests an oligoclonal activation. Thus, Hsps might serve as common activators for a range of clones of different antigen specificities (Kaufmann, 1996). So far, however, this response is specific to mice and γδ T cells responsive to Hsp do not appear to comprise a major subset in humans (Pfeffer *et al.*, 1990).

Functional studies have yet to establish a clear role for the γδ T cells. In mice, they occur in greater numbers in skin and the epithelial tissues of the intestine and female reproductive tract. The majority of γδ T cells in the mouse are not positively selected by MHC or even MHC-like molecules and there is some evidence that they may bind antigen on the surface of host cells directly. As appears to be the case with DN αβ T cells, it seems clear that the γδ T cells fulfil roles that fall outside of the classic model of the acquired immune response. Studies finding unusually high numbers of γδ T cell clones in the synovial fluid of arthritic joints suggest that their function cannot be ignored if we are to improve our understanding of autoimmune diseases. Studies at the cellular level are gradually providing clues to the function of γδ T cells, but as an alternative approach it may also prove useful to consider genetic and evolutionary data.

1.6 Structure and Organization of the Mouse TCR γ Locus

The genomic loci encoding the immunoglobulin and T cell receptor genes of mammals typically display a similar structural motif (Figure 1.4). They contain a relatively large number of V gene segments, followed by a smaller number of D segments in some genes and a smaller number of J segments followed by one or two C gene segments. Whereas the human TCR γ locus follows this motif, the murine TCR γ locus is arranged

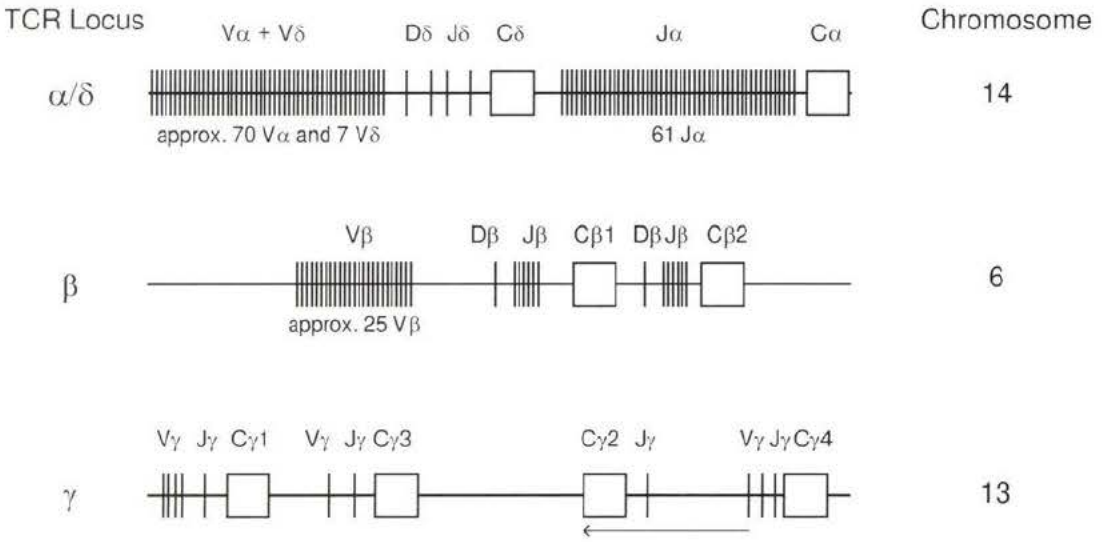

Figure 1.4. The organization of the mouse TCR gene loci. The number and location of variable (V), diversity (D), and joining (J) gene segments and constant (C) genes are shown. Arrow indicates transcription from opposite strand.

differently. The organization of the mouse TCRγ locus most closely resembles the mouse Ig λ light chain and the shark IgH loci. The mouse λ chain has two C regions, each with two J gene segments and one or two V gene segments (Lai *et al.*, 1989), and the shark has many clusters of a single V, J, one or more Ds and a C repeated throughout the genome (Kokubu *et al.*, 1988). The TCR α/δ locus is also unusual in that the locus encoding the TCR δ subunit is contained entirely within the TCR α locus, but the gene segments of the α/δ locus still follow the typical organization. The following section describes the physical structure of the TCR γ locus as we understand it to date.

The mouse TCR γ locus contains 7 V gene segments, 4 J gene segments and 4 constant region genes. These gene segments are arranged into four clusters of V, J and C gene segments. The first cluster, γ1, lies at the 5' end of the locus and contains 4 V gene segments, Vγ5, Vγ2, Vγ4 and Vγ3, followed by a single J (Jγ1) and a single C (Cγ1) (Figure 1.5) (Raulet, 1989). An enhancer element is located 2.5 kilobases (kb) 3' of the polyadenylation site of Cγ1 (Spencer *et al.*, 1991; Vernooij *et al.*, 1993). The γ1 cluster spans 41 kb with the V segments located within a 15 kb region, 17 kb upstream of Jγ1 (Vernooij *et al.*, 1993). The next cluster downstream is the γ3 cluster. It contains a single V, Vγ1.3, a single J, Jγ3 and a single C (Cγ3) (Raulet, 1989). The γ3 cluster spans 29 kb with Vγ1.3 located 13 kb upstream of Jγ3 (Vernooij *et al.*, 1993). The γ2 cluster lies further downstream in the opposite orientation to the other three clusters and consists of a single J (Jγ2) and a single C gene (Cγ2). Vγ1.2, the first of two V genes near the γ4 cluster, lies in the opposite orientation to Vγ1.1 and is only found transcribed with Jγ2 and Cγ2. Therefore Vγ1.2 can be considered to belong to the γ2 cluster. Including Vγ1.2, the γ2 cluster covers 36 kb with a V-J separation of 22 kb. The last cluster, γ4, has one V (Vγ1.1), a single J (Jγ4) and a single C (Cγ4) (Raulet, 1989). The γ4 cluster covers just 13 kb and has a V-J separation of just 2.2 kb (Vernooij *et al.*, 1993). V genes typically rearrange with the C in the same cluster though in rare cases Cγ4 has been found

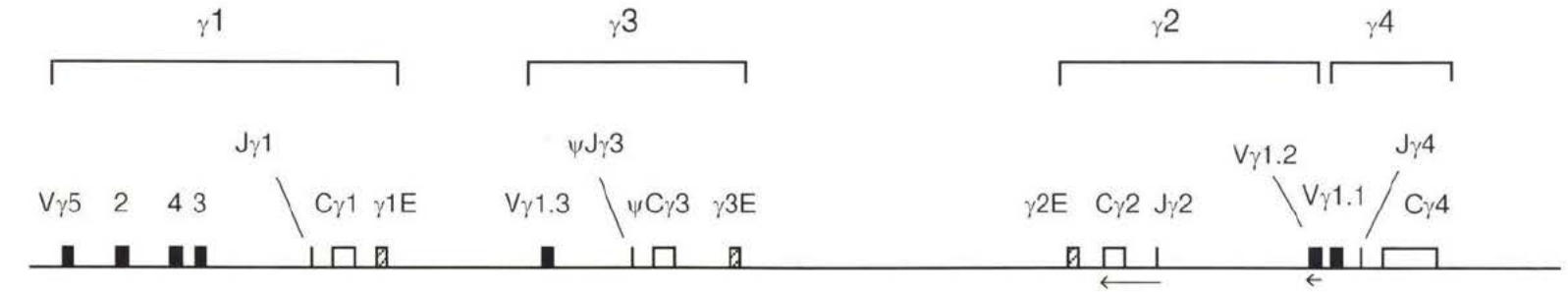

Figure 1.5. Scale map of the murine TCR γ locus. The TCR γ locus consists of four gene clusters (γ1, γ3, γ2 and γ4). Variable (V) and joining (J) gene segments are shown in black, constant (C) genes are shown as open boxes and enhancers (E) are shown as hatched boxes. Pseudogenes are indicated by the greek symbol, ψ. Arrows indicate direction of transcription. Adapted from Vernooij *et al.* (1993).

rearranged with Vγ5. These inter-cluster rearrangements are however nonfunctional (Raulet, 1989).

In the physical mapping of regulatory regions of the mouse TCRγ locus, Vernooij *et al.* (1993) located two putative transcriptional enhancers. A putative enhancer was located 3' of both Cγ2 and Cγ3 and exhibited 96% nucleotide sequence homology to the γ1 enhancer (Vernooij *et al.*, 1993). Interestingly, no enhancer-like sequence was detected in the γ4 cluster .

The Cγ1, 2 and 3 genes of the mouse are each encoded by 3 exons. The first exon is 330 base pairs (bp) in length and encodes the extracellular domain. The second exon (30-45 bp) encodes most of the hinge region and the third exon, 140 bp in length, codes for a transmembrane region and the cytoplasmic domain. The untranslated region ranges from 398 bp to 406 bp for Cγ1, 2 and 3 and >283 bp for Cγ4 (Hayday *et al.*, 1985; Garman *et al.*, 1986; Iwamoto *et al.*, 1986). Cγ1, 2 and 3 are over 90% homologous at the nucleotide level (Garman *et al.*, 1986) but Cγ3 contains a defective splice site which renders it nonfunctional (Hayday *et al.*, 1986). Jγ3 is also nonfunctional due to a stop codon in the coding sequence (Traunecker *et al.*, 1986). Studies in several mouse strains have failed to find any expression of these gene segments and the associated V gene (Vγ1.3) (Iwamoto *et al.*, 1986; Klotz *et al.*, 1989). One difference between Cγ1 and Cγ2 is that Cγ2 is missing an N glycosylation site which is present in the first exon of Cγ1. Cγ4 is the most divergent of the four C genes with just 66% homology at the nucleotide level (Garman *et al.*, 1986). The differences occur primarily in the length and sequence composition of the hinge and transmembrane regions. Unlike the other three C regions, the Cγ4 gene consists of four exons. The second and third exons encode 18 and 15 amino acids of the hinge region. Although the γ4 cluster is the smallest of the four, the Cγ4 gene is the largest of the four C genes, spanning nearly 8 kb, compared with 2.5 kb for Cγ1, 2 and 3 (Vernooij *et al.*, 1993). The homology seen in the C regions holds for the J gene

segments as well, with Jγ4 being the most diverse (Garman *et al.*, 1986). The 5' recombinase recognition sequences of the J gene segments are all spaced by 12 bp (Hayday *et al.*, 1985; Traunecker *et al.*, 1986).

Of the seven Vγ gene segments, Vγ1.1, Vγ1.2 and Vγ1.3 are the most homologous with >94% nucleotide similarity and make up the Vγ1 family. Of the remaining four V gene segments in the γ1 cluster, Vγ2 shows the most similarity to the Vγ1 family (65-70% at the nucleotide level). The other three Vγ segments are more divergent and show little similarity to any other Vγ gene segments in the locus. Recognition sequences 3' of the V genes are spaced by 23 bp (Hayday *et al.*, 1986; Garman *et al.*, 1986). Vγ1.1 and Vγ1.2 lie side by side near Jγ4 and Cγ4. Vγ1.2 lies inverted, 5' of Vγ1.1 so that they are closely opposed in a head to head arrangement (Hayday *et al.*, 1986).

The unique organization of the mouse TCR γ locus makes it an interesting target for comparative evolutionary studies. Comparisons of DNA sequence data for the locus may provide clues to the origin of its unique organization, put a time to the duplication events that led to the four clusters, and identify conserved sequences which may indicate some function.

1.7 Evolution of the Acquired Immune Response

The TCR and Ig are related molecules that share a remarkable system for generating diversity. The role of these receptors in immunity is a critical one and there has been a great deal of interest in determining how such a system might evolve. Past studies of the evolution of the immune system have used immunological techniques, such as the level of cross-reactivity between molecules, to determine relatedness. In recent years, however, the technology has become available to obtain amino acid and nucleotide sequences from which much more detailed comparisons can be made. Using these techniques it is possible to find related molecules in a variety of species. These molecules can then be compared to better

understand the evolutionary pathways that shaped the immune systems of present vertebrate species.

Immunoglobulins have been found in all placoderm-derived orders from elasmobranchs to mammals (Litman *et al.*, 1971; Marchalonis, 1977). No Ig, TCR, or MHC proteins have been found in jawless vertebrates. In all species which have been found to express γδ TCRs there is evidence for the expression of multiple forms, or isotypes, of the Cγ gene. Humans express two forms, mice express three forms, cows express at least three forms and sheep express five forms (Hein and Dudler, 1993). These different isotypes vary in sequence and structure within species. The extracellular N-terminal domains of the human C region isotypes are almost identical, differing at just 3 amino acid residues (Lefranc *et al.*, 1986; Krangel et al, 1987; Garman *et al.*, 1986; Hayday *et al.*, 1986). The same region of the Cγ1 and Cγ2 isotypes of the mouse are also virtually identical at the amino acid level, but Cγ4 differs by 18 residues (Iwamoto *et al.*, 1986). The Cγ isotypes of the artiodactyls (sheep, cow, and pig), with 29 to 48 amino acid residues difference, are as different from each other as they are between orthologues (Hein and Dudler, 1993). This suggests that the Cγ isotypes of the artiodactyls are the products of relatively ancient duplication events that occurred in a common ancestor, while the isotypes of the human and mouse probably arose after the divergence of the human and rodent lineages (Hein and Dudler, 1993). A phylogenetic analysis of all available cDNA sequences of TCR Cγ genes from humans, mice and artiodactyls supports this interpretation (Ciccarese *et al.*,1997).

The greatest divergence between C genes occurs in the hinge region. The hinge region is encoded by one or more exons each encoding a 16 amino acid unit. Studies of human Cγ1 and Cγ2 suggest that diversity in the hinge region is the result of a duplication or triplication of an exon encoding this 16 amino acid unit (Lefranc *et al.*, 1986; Krangel et al, 1987). The repetition of 16 amino acid units in the hinge region of sheep, cow and pig

Cγ chains suggest that these duplication or triplication events occurred in these lineages as well (Hein *et al.*, 1990; Takeuchi *et al.*, 1992; Thome *et al.*, 1993). In contrast to the Cγ genes, Cδ genes are relatively highly conserved between mammalian species and there is no species known to contain more than a single Cδ isotype (Hein *et al.*, 1990; Takeuchi *et al.*, 1992; Thome *et al.*, 1993; Loh *et al.*, 1987; Chien *et al.*, 1987).

Of the TCRs, the TCR α subunit shows the greatest variability in the extracellular domain and numerous insertions and deletions can be seen between species (Hein and Dudler, 1993). The Cβ chains on the other hand are the most conserved of all the TCR subunits with 77% to 94% identity in the amino acid sequence between different mammalian species (Thome *et al.*, 1993; Ito *et al.*, 1993; Yanagi *et al.*, 1984; Jones *et al.*, 1985; Hedrick *et al.*, 1984; Malissen *et al.*, 1984; Williams *et al.*, 1991; Angiolillo *et al.*, 1985; Marche and Kindt, 1986; Grossberger *et al.*, 1993; Tanaka *et al.*, 1990).

In humans and mice there is limited combinatorial diversity in the γδ TCR due to a relatively low number of γ and δ gene segments. In artiodactyls this repertoire is much higher. At least 13 Vγ and 28 Vδ gene segments have been found in sheep versus 6 Vγ and 7 Vδ gene segments in the mouse and 8 Vγ and 3 Vδ segments in human (Hein and Dudler, 1993; Ishiguro *et al.*, 1992). Like the Igλ light chain V regions, which are similar in their limited diversity and frequency (Hein and Dudler, 1993), the mouse TCR Vδ gene products exhibit high variability in the CDR 1 and 2 regions (Schiffer *et al.*, 1986; Schiffer *et al.*, 1992; Bougueleret and Claverie, 1987). The CDR3 region of the TCR Vδ shows less variability than the CDR3 of immunoglobulins, but this is offset by the somatic contribution of the Dδ gene segments, two of which can be used in a single receptor (Hein, 1994). These similarities between the Igλ V region and the δ V regions suggest a relatedness that may extend to function (Schild *et al.*, 1994).

As in Igs, the CDR3 of TCRs, which occurs at the VJ and VDJ junctions, is responsible for a large part of the variation in the receptor due to the random action of

exonucleases and the insertion of nucleotides between recombining gene segments. The γδ TCRs differ from αβ TCRs in that the δ chain CDR3 is significantly longer than that of the γ chain gene (Rock *et al.*, 1994). Additionally, the length of the CDR3 is substantially more variable in TCR δ than in the other TCR subunits (Rock *et al.*, 1994). The large CDR3 of the TCR β and δ subunits lend a tremendous amount of variability to their V regions, yet the C regions of these subunits are the most highly conserved of the TCR subunits. In the TCR β and δ chains there seems to have been a selection for diversification of the antigen binding region, but a conservation of some property associated with the constant region. This is similar to a situation involving IgM, which also has a highly conserved constant region. During B cell ontogeny the IgM heavy chain is paired with a surrogate light chain and is expressed on the cell surface where it may mediate early selection events (Melchers *et al.*, 1993). There is some evidence that a similar phenomenon may occur with the TCR δ and a surrogate α chain (Groettrup *et al.*, 1993).

It is clear through sequence homology, structural homology and their shared genetic organization that Igs and TCRs share a common ancestry. This similarity is particularly evident in the γδ TCRs. The existence of multiple, structurally divergent Cγ genes throughout the artiodactyls suggests an ancestral pool of structurally divergent Cγ genes in early γδ TCRs, and the 5 diverse Cγ chains of sheep are similar in number and diversity to IgH chains. The similarities between the γδ TCRs and Igs suggest either that the two loci diverged from a common ancestor or that the two loci underwent parallel evolution after their initial divergence (Hein, 1994).

The current model of T cell antigen recognition requires that antigen be presented to T cells in the context of an antigen presenting molecule. The TCR then recognizes both the presenting molecule and the processed antigenic peptide located in a peptide binding groove of the presenting molecule (Germain, 1994). Thus the origin of T cells would have required the association with a primordial TCR with a primordial antigen presenting

molecule. Since Ig and MHC molecules have been described in all vertebrates studied from elasmobranchs to mammals but have not been found in jawless fish, the adaptive immune response is thought to have evolved at least 450 million years ago (Kasahara *et al.*, 1992).

A class I MHC orthologue isolated from the amphibian *Xenopus* showed similarity in the peptide binding region to members of the Hsp70 family (Flajnik *et al.*, 1991a). Produced in response to cell stress, Hsp70 expression has been shown to be induced in response to injection of denatured protein (Pelham, 1988). It has been suggested that the Hsp70 family may act as chaperones involved with protein folding or escorting proteins across membranes (Flajnik, 1994). Hsp70 has been shown to bind small peptides (Flynn *et al.*, 1989). The similarity between *Xenopus* class I MHC and Hsp70 is restricted to a putative peptide binding cleft. Comparison of the secondary and tertiary structure of the two proteins supported the idea of homology between Hsp70 and MHC class I molecules (Flajnik *et al.*, 1991b), however recent crystallographic data has shown clear differences in the petide binding clefts of Hsp70 and MHC class I (Zhu *et al.*, 1996).

In all of the vertebrates studied to date there are invariant MHC class I molecules called class Ib. Unlike classic MHC, class Ib molecules are under no selection for amino acid changes in the peptide binding cleft and may actually be under selection against change (Flajnik, 1994). These molecules are expressed in epithelial surfaces such as skin, gut, uterus and lung and bind conserved proteins derived from the heat shock proteins of pathogens. The class Ib molecules are recognized by T cells with very restricted TCR repertoires (Janeway, 1992).

From this evidence and a model put forward by Janeway (1992), Flajnik (1994) has proposed that the adaptive immune response evolved at the epithelia. Primitive T lymphocytes possessing restricted repertoires would have recognized conserved chaperone proteins bound to monomorphic presentation molecules on the surface of stressed epithelial cells. These molecules would then have coevolved with the low-diversity TCRs,

increasing in diversity and allowing the development of an acquired cellular immune response and self-tolerance (Flajnik, 1994).

Hein (1994) proposed a model of lymphocyte evolution which is consistent with the theories of Janeway (1992) and Flajnik (1994). Hein's (1994) model places the origin of the γδ T cells at the epithelia, evolving with an as yet unidentified class of antigen presenting molecule which may still be involved in γδ TCR antigen recognition. The γδ T cells and B cells would have separated from a common lymphoid precursor which recognized archaic ligands of low diversity presented by epithelial cells during lymphopoiesis (Hein, 1994). While mature B cells functioned independently of this ligand, mature γδ T cells would have maintained a requirement for these molecules in antigen recognition (Hein, 1994). A subclass of these ligands may have acquired the capability of binding small peptide fragments and become expressed more widely in the body. The development of specialized antigen presenting cells capable of sampling both their internal and external environment would have involved the evolution of polymorphic MHC molecules and a specialized class of TCR which would give rise to the αβ TCR (Hein, 1994). Hein (1994) called his theory "MHC capture", referring to the capture of MHC molecules by pre-existing antigen receptors as opposed to the "escape" hypothesis of Davis and Bjorkman (1989), which suggests that the Ig molecules escaped MHC by acquiring the ability to bind free antigen.

An earlier hypothesis by Matsunaga and Dahl (1989) suggested that the γδ T cell receptor is an evolutionary relic representing the primitive immune system, predating both Ig and αβ TCR. Organization of the TCR α/δ locus suggests that TCR δ was the precursor of TCRα. Duplication of the TCRδ gene segments from the centre of the locus would then have created the α gene segments. The increasing diversity of the B cell V repertoire would have resulted in lower clonal populations and necessitated the coordination of the immune response by the T helper system. To meet the new requirement for a helper cell, the αβ T

cell subset would have grown from a subset of T cells that was able to co-evolve with a diversifying MHC repertoire. The γδ TCR is relatively invariant in the CDR1 and CDR2 regions, which are thought to recognize MHC, but has large CDR3 diversity, suggesting that antigen recognition is important in the γδ TCR while the antigen presenting molecule remains invariant or is perhaps not required. This model of T cell evolution suggests that cytotoxic γδ TCRs may predate T helper function, while the MHC class I restricted cytotoxic αβ T cell response could have arisen with the divergence of a CD4/8 precursor in parallel with the evolution of MHC class I and II (Matsunaga and Dahl, 1989).

Putting our current knowledge of γδ TCR structure and function into an evolutionary context is a useful exercise in establishing the role of γδ T cells in the immune system. Current knowledge, which is still very sparse, suggests that the γδ TCR may predate the αβ TCR in evolution. Matsunaga and Dahl (1989) argue that their current role in the immune system is a primitive one. While this may be, the continued presence of γδ T cells in higher vertebrates suggests that their role continues to contribute some degree of fitness and has not become redundant. Interestingly, while the γ and δ loci of the human and mouse retain ancestral state in having low numbers of gene segments, the γδ TCRs of artiodactyls have evolved a much larger number of gene segments. Thus, it appears as though the γδ TCRs in the human-rodent and the artiodactyl lineages have followed different evolutionary paths. Attempting to understand the past events which have shaped these paths will be important in reaching an understanding of the current role of the γδ T cells.

1.8 Origin of the Mouse TCRγ Locus

In the genomic DNA a relatively small number of genes are interspersed amid a sea of non-coding sequence. While the genes are generally conserved, non-functional intervening DNA is under no selection pressure and should acquire mutations at a much

higher rate. Thus, in pairwise comparisons of genomic DNA sequences, conservation of non-coding DNA suggests either that the conserved region serves some function or that it is the product of a recent duplication event.

Differences in the physical maps of the TCR γ loci of human (LeFranc *et al.*, 1990) and mouse (Hayday, 1986; Raulet, 1986; Garman, 1989; Vernooij *et al.*, 1994) suggest that the TCR γ locus has undergone substantial change since the primate-rodent split 80 million years ago (MYA). The high homology between mouse Cγ1, 2 and 3 suggest that they are the products of a recent duplication events while the more divergent Cγ4 is the product of a more ancient duplication (Garman *et al.*, 1986). The organization of the gene segments into four distinct clusters of V, J and C segments suggests that the duplications involve this basic unit of three gene segments. However, the Vγ1.1 and Vγ1.2 genes, which belong to the γ2 and γ4 clusters respectively, are over 95% similar, despite the low homology of their respective C genes. Vernooij and co-workers (1993) suggest that the γ4 cluster was first duplicated from a region containing a single V and the Jγ1 and Cγ1 elements. Clusters γ2 and γ3 would have been duplicated more recently, with the γ2 cluster originating with an inverted duplication of Jγ1 and Cγ1 and Vγ1.2 having formed separately from an inverted duplication of the Vγ1.1 segment. The γ2 cluster would then have been duplicated and inverted to form the γ3 cluster. The large distance between Jγ2 and Vγ1.2 might have resulted from a later insertion of a retroviral sequence between the two gene segments (Vernooij *et al.*, 1993).

In order to study the evolutionary events occurring within the mouse TCR γ locus we obtained the continuous genomic DNA sequence of the 3' end of the mouse TCR γ locus containing the γ2 and γ4 clusters. Comparative analysis of this data was used to investigate the extent of internal duplications in the region and the relative time frame in which they have occurred.

MATERIALS AND METHODS

2.1 Cosmid Library

A cosmid (pWE15a) library covering the mouse TCR γ locus was provided by Dr. Bernard Vernooij (Vernooij, 1994). Three overlapping cosmids, designated Mγ84, Mγ3 and Mγ74, spanning the γ2 and γ4 clusters at the 3' end of the locus were selected from this library and used to determine the genomic sequence.

2.2 Cosmid Preparation

A single isolate of each cosmid clone was grown overnight in Luria Broth (LB) + Amp [40μg/ml] and the cosmid DNA was extracted by large scale alkali lysis (Sambrook *et al.*, 1989). The cosmid DNA was digested with *Eco*R1 and *Hin*dIII and restriction fragment sizes were checked against known values where possible. The size of the insert was estimated by totalling the restriction fragment sizes for each cosmid digest. Restriction digests were also checked for background smearing. The presence of background smearing was interpreted as evidence of contamination with *Escherichia coli* genomic DNA.

Cosmids were sequenced using a shotgun approach. This approach involves breaking the target DNA into random fragments and then ligating those fragments into a sequencing vector in a single ligation reaction. This provides a sequencing library of cloned fragments which covers the entire length of the target several times over. Subcloned inserts were then sequenced and assembled with the aid of computer software to yield a highly accurate contiguous sequence of the target DNA.

2.3 Sequencing Library Construction

2.3.1 Fragmentation of Cosmid DNA

Following the protocol of Bodenteich *et al.* (1994), 50 μg of cosmid DNA was suspended in 2.0 mL of tris-magnesium-glycerol buffer (50 mM tris-HCl, pH 8.0 containing 15 mM $MgCl_2$ and 10% glycerol). This DNA solution was then added to the

the instructions of Bodenteich and co-workers (1994). The nebulizer was placed in an ice water bath and nebulization was performed at 30 psi (N_2 gas) for 150 seconds. The nebulizer was then centrifuged at 1000 x g for 1 minute to recover as much of the DNA solution as possible from the walls of the nebulizer. Sheared DNA was recovered from solution by ethanol precipitation and resuspended in 89 µl of TE buffer (10 mM Tris · HCl (pH 8.0), 1 mM EDTA (pH 8.0)).

2.3.2 Insert Preparation

The nebulized fragments were made blunt ended according to the method of Povenelli and Gibbs (1993). The entire 50 µg sample of nebulized insert DNA was incubated with 50 U of mung bean nuclease in a 100 µl reaction volume of 1x mung bean buffer (30 mM sodium acetate, pH 4.6, 50 mM $NaCl_2$, 1.0 mM zinc acetate, 0.001% Triton X-100, Boehringer) at 37°C for 30 min. DNA was then recovered by 1:1 phenol:chloroform extraction and precipitation with 0.5 volumes of 7 M ammonium acetate and 2 volumes of ethanol. The DNA pellet was then resuspended in 20 µl TE buffer. The nuclease treated fragments were then incubated with 100µl of each dNTP, 1x Boehringer buffer "B" (10.0 mM Tris-HCl, pH 8.0, 5.0 mM $MgCl_2$, 100.0 mM $NaCl_2$, 1.0 mM 2-mercaptoethanol) and 40 U of T4 DNA polymerase (Boehringer) in a reaction volume of 400 µl for 10 min at room temperature (RT). Twenty-five units of Klenow (Boehringer, sequencing grade) was added to the reaction and incubated for 10 min at RT before moving the reaction to 16°C for 2 hours. DNA was then extracted as described above and resuspended in 90 µl TE.

Blunt fragments were then size fractionated by gel electrophoresis as follows. Fragments were loaded into a 5 cm wide well on a 2% agarose gel. Bacteriophage λ genomic DNA cut with *Hin*dIII (λ/*Hin*dIII)was used as a size marker and run in lanes neighbouring the blunt ended fragments. A size marker lane was then excised from the gel, stained with ethidium bromide and marks were made at the 1.5 kb to 2 kb range under UV

light. The marked lane was then returned to the gel and a slice of the sample lane corresponding to the marked size range was carefully excised. The gel was then cut in half and the top portion, from which the size fraction was removed, was photographed under UV light while the bottom half was used for support in the steps to follow. A notch approximately 1 cm wide and 1.5 cm deep was cut in the gel. At the bottom of the notch a small piece of activated DEAE paper was placed (Sambrook *et al.*, 1989). The cut fraction of gel containing the blunt fragments of interest was then cut into smaller pieces and these pieces were placed into the notch against the DEAE paper. The blunt fragments were electrophoresed onto the DEAE paper. The DEAE paper was then removed and the DNA was eluted in high-salt buffer (Sambrook *et al.*, 1989). The eluted DNA was precipitated with ethanol and resuspended in 10 µl dH_2O.

2.3.3 Vector Preparation and Ligation

Vector was prepared by digesting 5 µg of M13mp19 RF DNA with 80 U of *Hinc*II in a total volume of 50 µl, and incubated at 37°C for 1.5 hours. The restriction digest was then extracted with 1:1 phenol:chloroform and precipitated with 0.1 volume of 3M sodium acetate and 2 volumes of 95% ethanol. DNA was collected by centrifugation in a microfuge for 30 minutes at high speed. The supernatant was then decanted and the pellet was washed in 70% ethanol and centrifuged for 5 minutes. The supernatant was again decanted and the pellet allowed to air dry for 10 minutes before resuspending in 88 µl of dH_20. The cut vector was then dephosphorylated with calf intestinal phosphatase (CIP) in 1x CIP buffer and 2 U of CIP in a reaction volume of 100 µl and incubating 1 hour at 37°C. The dephosphorylation reaction was stopped by adding 0.1 volume of 200 mM EGTA and heat denaturing at 65°C for 10 minutes. Two phenol:chloroform extractions were then performed, and the sample was ethanol precipitated and resuspended in 100 µl of 0.5x TE (pH 8) (Sambrook *et al.*, 1989).

To estimate the concentration of the size fractionated insert and to check that the vector had been digested to completion, 2 µl of each was electrophoresed on a 0.7% agarose gel alongside a measured amount of λ/*Bst*E ladder. Ligations were performed with three different ratios of vector to insert, using a ligation of vector with no insert added as a control. The first ligation mixture contained a ratio of 1 vector to 0.5 insert, the second contained 1 vector to 1.5 insert, and the third contained 1 vector to 4 insert. Each reaction used 50 ng of vector, 0.5 U of T4 Ligase in a total volume of 15 µl in 1x blunt-end ligation buffer. Ligations were incubated overnight at 16°C.

2.3.4 Cloning and Plating

Ligation reactions were either transformed into high-efficiency competent *E. coli* DH5α (Gibco BRL) or electroporated into the same bacterial strain with a Bio-Rad Gene Pulser (Bio-Rad Laboratories Ltd.). The transformed or electroporated cells were then plated with lawn cells in yeast-tryptone (YT) top agar containing X-Gal and IPTG and incubated overnight at 37°C.

Colourless plaques were picked with sterile toothpicks and used to inoculate tubes of LB or YT broth containing DH5α host cells. After an overnight incubation, the bacterial cells were pelleted, leaving the M13 phage particles in the supernatant. Template DNA was purified using PEG precipitation. Growth tubes were centrifuged at 3000 g to pellet bacterial cells and cell debris and approximately 1.2 ml of supernatant was decanted from each tube into a clean 1.5 ml microfuge tube. Phage particles were precipitated from the supernatant by adding 250 µl of PEG/$NaCl_2$ (20% PEG 8000, 2.5 M $NaCl_2$) and centrifuging at approximately 14000 g for 30 min. Pellets were then resuspended in 100 µl TE and extracted with phenol, 1:1 phenol:chloroform and chloroform. Template DNA was then precipitated with 8 µl 3 M sodium acetate and 150 µl 95% ethanol. Alternatively, some template preps were performed with QIAprep 96 M13 (Qiagen Inc.) column purification kits.

2.4 Sequencing and Assembly

Purified template DNA was sequenced by the Sanger dideoxy method on an ABI 373 Stretch semi-automated DNA sequencer with ABI Prism sequence reaction kits (Perkin-Elmer Corp.).

After obtaining approximately 500 sequences, sequence data files were assembled with the aid of either SeqMan (DNASTAR Inc.) on the Macintosh platform or Phred, Phrap and Consed (Phil Green *et al.*, unpublished data. E-mail: phg@u.washington.edu) on a UNIX platform. New sequences were added to the assembly until a single contiguous sequence for the target was obtained. Once a contiguous sequence was obtained it was checked manually and edited to correct obvious errors. Minimal coverage was considered to be two sequences in one direction and one in another. Where these requirements were not met additional sequencing of clones in that region was performed until the minimal coverage was obtained. If this coverage could not be obtained, resequencing of a single clone with dye-linked terminators rather than dye-linked primers was performed.

2.5 Gap Filling

According to the physical map of the region (Vernooij *et al.*, 1994), the ends of the two cosmids sequenced were approximately 5 kb apart. In order to fill this gap in our genomic sequence we used the sequences obtained from the ends of the two finished cosmids to design primers for PCR. The primers used were named Gap1 (5'-TGCCTTGTTCCACGTCAGTAA-3') and Gap2 (5'-CAGAGTAATTGGGGCTTCA-3'). The long PCR protocol of Cheng and co-workers (1994), modified by workers at the George Church lab (http://twod.med.harvard.edu/labgc/estep/longPCR_protocol.html), was used to amplify a fragment from the overlapping cosmid, Mγ74. The same protocol was used to check for the presence of the region 5' of the first exon of the Cγ2 gene in the cosmid Mγ3 stock cultures. A PCR primer was designed from a known sequence at the end of the third exon of Cγ2. This primer, called Cγ2 UTR (5'-

ACGAGTGGTGGTACAGCAAGTCAG-3'), and the Enh96 (5'-AAGCTTAGGGATATTTTGTCTC-3') primer (Vernooij *et al.*, 1994) were used to amplify the region from the end of the Cγ2 gene to the end of the γ2 enhancer. Following verification of the template, another primer, Cγ2 Exon I (5'-AGCCCACTATTTTTCTTCCTTCTG-3'), was used with the Enh96 primer to obtain a PCR product that would extend from the start of the first exon of Cγ2 to the end of the γ2 enhancer.

For both gaps, approximately 190 PCR reactions were pooled then subjected to the shotgun cloning and assembly protocols described previously. Contigs and gaps were then aligned manually to form a single contiguous sequence.

2.6 Sequence Analysis

The final sequence was analyzed for internal homology by a dot matrix plot against itself and its complement using Inherit (ABI, Perkin-Elmer Corp.) and DOTTER (ftp://sanger.ac.uk/pub/dotter, Sonnhammer and Durbin, 1996). Inherit plots were generated using a 40 bp window, and a 20 bp shift. DOTTER plots were generated using the default parameters (25 bp window).

Pairwise alignments of homologous regions were performed using SIM (Huang and Miller, 1991) with the default parameters. By default SIM sets the cost of a matching aligned pair equal to 1 (M=1), the cost of a transition equal to -1, the cost of a transversion equal to 1, the gap opening penalty equal to 6.0, and the gap extension penalty equal to 0.2.

The genomic sequence was searched against a database of rodent repetitive elements using Repeat Masker (http://ftp.genome.washington.edu/cgi-bin/mrs/mrs_req, Smith and Green, unpublished data). The sequence file was sent via the internet to the Repeat Masker site and the type of sequence was specified as rodent. Results of the search were returned via the internet.

TCR γ cDNA sequences utilizing Cγ2 and Cγ4 isotypes were obtained from Genbank and compared to the genomic sequence by dot matrix plot using DOTTER. Lines of identity between the cDNA and the genomic sequences indicated the locations of the gene segments and exons. The cDNA sequences were then broken down into gene segments and exons. These sequences were aligned with the genomic sequence in SeqMan to provide a sequence level reference to the gene segments, exons and the flanking regions.

In order to examine the consequences of VJ recombination, 12 cDNA sequences for Vγ1.1 and Vγ1.2 rearrangements were recovered from GenBank using the string search feature of Inherit. A short sequence from each V and J gene segment was used to perform boolean searches to retrieve cDNA sequences containing both a Vγ1.1 and a Jγ4 gene segment or a Vγ1.2 and a Jγ2 gene segment. The retrieved cDNA sequences were then aligned with the germline sequences adjacent to the recombination signals to allow the characterization of the nucleotide deletion and addition events occurring at the VJ junctions.

The sequence was searched for unmapped exons using XGRAIL 1.2 (Uberbacher and Mural, 1991), BLAST (NCBI), and BLAST/BEAUTY (Baylor College of Medicine Search Launcher). XGRAIL uses a complex pattern recognition algorithm to identify sequences in genomic DNA that exhibit characteristics of exons. Thus, XGRAIL was used to locate exons that might not be represented by full length cDNA or protein sequences. XGRAIL was also used to locate repeats, CpG islands and GC content. BLAST is a fast search algorithm offered by NCBI to search DNA or protein sequence database for sequences matching a user supplied query sequence. BEAUTY is a front-end to BLAST which processes a nucleotide query sequence, translating it in all six reading frames, then submits the resulting protein sequences to BLAST.

Restriction sites in the region were located with MapDraw (DNASTAR Inc.). Percent GC and number of CpG sites were calculated over an 800 bp sliding window with 600 bp shifts using in-house software (Koop, unpublished).

2.7 Phylogenetic Analysis

cDNA and amino acid sequences were obtained from Genbank through the NCBI Entrez Nucleotide Query (http://atlas.nlm.nih.gov:5700/Entrez/index.html). The protein sequences of TCR Cγ genes from human (Lefranc *et al.*, 1986), mouse (Hayday *et al.*, 1985), rat (Kim *et al.*, 1994; Kinebuchi *et al.*, 1994), rabbit (Isono *et al.*, 1995), sheep (Hein and Dudler, 1993), cow (Takeuchi *et al.*, 1992; Ishiguro *et al.*, 1993), pig (Thome *et al.*, 1993), and minke whale (Zhang *et al.*, 1995) were aligned using CLUSTALW at the Baylor College of Medicine Search Launcher. The CLUSTALW alignment was edited using Se-Al (Andrew Rambaut, http://evolve.zoo.ox.ac.uk/Se-Al/SeAl.html). A parsimony tree was then constructed from the protein sequence alignment using PAUP* (David Swofford, University of Madisson). A heuristic search algorithm was used to find the most parsimonious tree and branching patterns were tested for significance by non-parametric bootstrap analysis of 100 replicates. The length of the alignment was 255 characters and 136 characters were parsimony informative. Equal weight was given to all characters and starting trees were obtained by stepwise addition of taxa.

RESULTS

3.1 Large Scale Sequencing

To obtain the genomic sequence of γ2 and γ4 clusters of the mouse TCRγ locus, a shotgun sequencing strategy was used. From the physical map of the locus, two cosmid clones, γ3 and γ84, were chosen to provide coverage of the region of interest (Figure 3.1). Cosmid γ3 provided the sequence of the 5' end of the region and γ84 provided the 3' end. An overlapping cosmid, γ74, was used to fill a gap between the ends of γ3 and γ84 using a PCR based strategy. Stability problems were encountered with cosmid γ3, that resulted in the deletion of the 5' end of the insert including most of the Cγ2 gene and the γ2 enhancer. PCR was used to obtain the missing piece of the γ2 cluster and the resulting sequence was verified by restriction enzyme analysis, PCR, and comparison with cDNA and genomic sequences from Genbank.

3.1.1 Cosmid Sequencing and Assembly

A contiguous sequence, or contig, of 34 kb covering cosmid γ84 was assembled from 783 sub-clone sequences. The final contig contained 619 sequences and had an average 10.1 fold redundancy. 48 sequences were of the M13 sequencing vector and 35 sequences were of the cosmid vector. An additional 81 sequences fit into no major contig and probably represent contamination with *E. coli* genomic DNA.

Cosmid γ3 was finished after obtaining 813 sequences. On assembly, a 20.6 kb contig of 702 sequences with an average 19 fold redundancy was obtained. This was first thought to be an erroneous assembly, but it was found to be the only assembly possible. Cosmid vector was encountered at both ends of the contig, but at the 5' end of the sequence, the vector did not meet insert at the multiple cloning site, but at a site over halfway within the cloning vector. In addition, a BLAST search of the sequence immediately flanking the vector at the 5' end of the insert produced a perfect match with an insertion sequence (IS R) of an *E. coli Tn*10 transposable element (Figure 3.2A). The

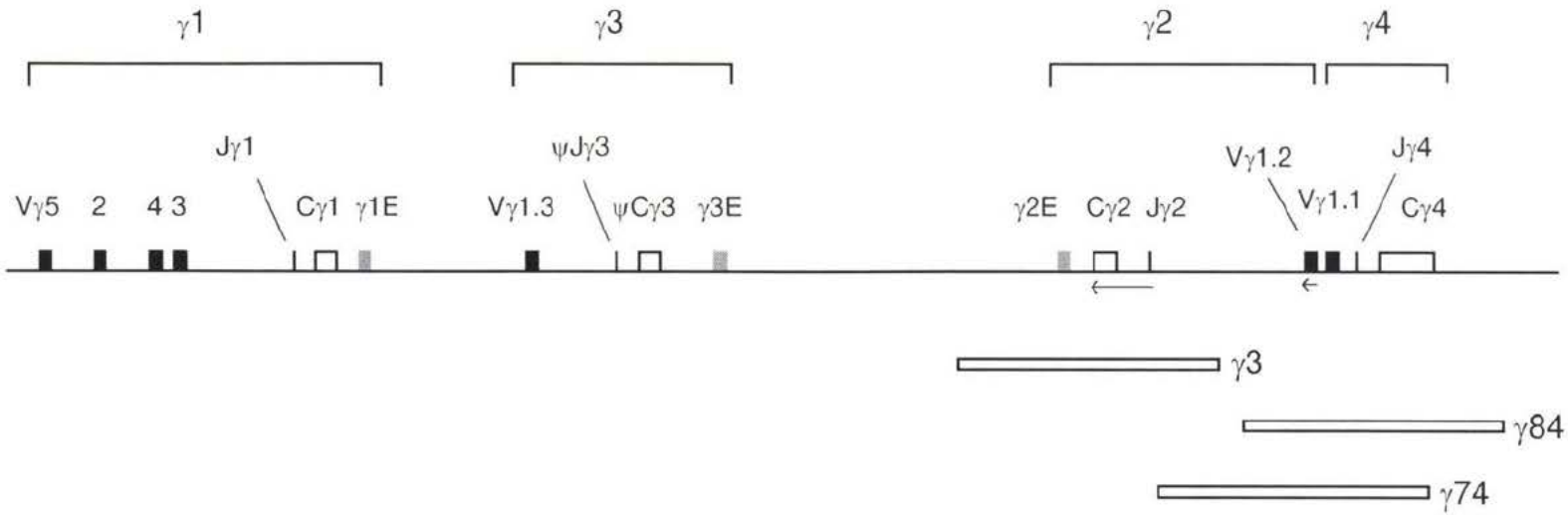

Figure 3.1. Map of the mouse TCR γ locus indicating the regions covered by the γ3, γ74 and γ84 cosmids. Arrows indicate direction of transcription. Adapted from Vernooij *et al.* (1993).

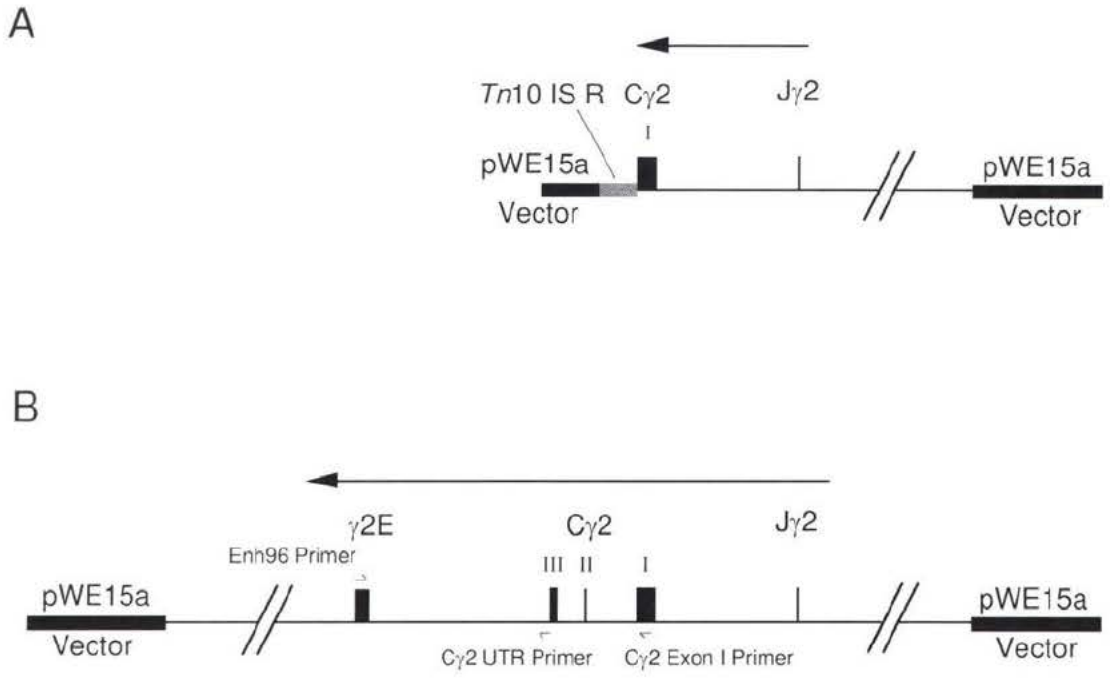

Figure 3.2. Comparison of rearranged and intact cosmid γ3 inserts. A. A rearranged γ3 cosmid containing the Jγ2 gene segment and exon 1 of the Cγ2 gene of the mouse TCRγ locus. A *Tn*10 IS R element was found adjacent to the 3' end of the Cγ2 exon. B. An intact γ3 cosmid. Primer sites indicated by half arrows were used to first establish the presence of the sequence missing from the rearranged clone (Enh96 primer to Cγ2 UTR primer), then to amplify a portion of the missing region from the first exon of Cγ2 to the end of the γ2 enhancer (Enh96 primer to Cγ2 Exon 1 primer). Arrows indicate direction of transcription. The insert is denoted by a thin line and vector (cosmid pWE15a) is indicated by a thick line.

sequence immediately 3' of the *Tn*10 IS indicated a 100% match with the first exon of the constant region Cγ2 which was expected to occur near the middle of cosmid γ3. It appeared that part of the vector together with the insert 5' of the first exon of Cγ2 had been deleted with the insertion of the *Tn*10 IS R element. This evidence forced us to conclude that cosmid γ3 had undergone a rearrangement and deletion during its growth in culture. Restriction enzyme analysis and PCR evidence suggested that the remaining insert had not been altered.

Attempts to isolate a non-rearranged clone were made by returning to the original stock cultures of the γ3 clone. Cosmid DNA extractions were performed on approximately 230 well isolated colonies, but restriction digests showed either the wrong pattern or no pattern at all (data not shown). This could suggest that all available cultures of γ3 were contaminated with a bacterial strain with no cosmid that greatly outnumbered those clones containing the correct cosmid. In order to test the possibility that clones harboring the original γ3 cosmid were present in these cultures, three stocks from the original source (Vernooij, 1993) were subjected to PCR. A swab of colonies of each clone was placed into buffered solution and a long PCR protocol using primers Enh96 (Vernooij, 1993) and Cγ2UTR which covered the region from the 3' end of the γ2 enhancer to the sequence following the stop codon of exon three of Cγ2 (Figure 3.2B). Two of the three stock cultures produced a single band of approximately 4.5 kb (Figure 3.3). A positive control of mouse genomic DNA produced two bands. The brightest was approximately 3.8 kb in length, but a faint band corresponded exactly to the 4.5 kb band of the γ3 clone stocks. The smaller, brighter band in the genomic DNA corresponded to the homologous region of the γ1 cluster, which was amplified due to its high homology (>96%) with Cγ2 and the γ2 enhancer (Vernooij *et al.*, 1993). Both bands agreed with the distances estimated in the physical map of the locus (Vernooij *et al.*, 1993). Thus it was established that some of

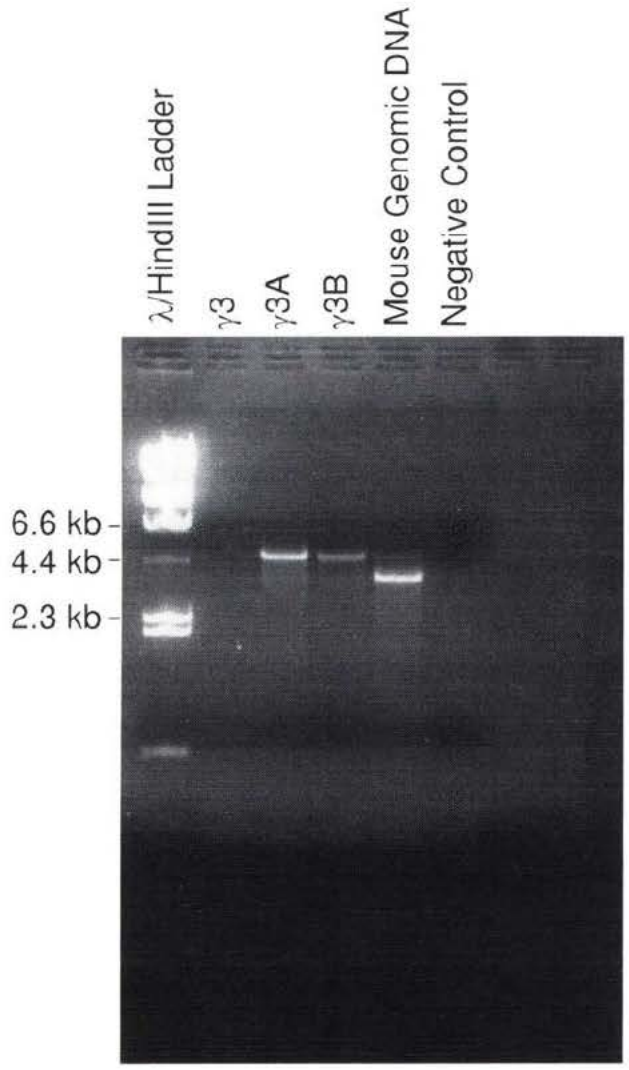

Figure 3.3. PCR amplification from the end of the γ2 enhancer to the start of the Cγ2 3' untranslated region. Lanes γ3, γ3A, and γ3B represent amplification from colony swabs of the original clones γ3, γ3A and γ3B respectively. Mouse genomic DNA was used as a positive control. Two amplification products are present in the positive control. The bright, lower molecular weight band is the product of amplification of the homologous region of Cγ1, while a band of lower intensity corresponds with bands in the sample lanes.

cells from the original clone stocks harboured a section of the γ3 cosmid containing the Cγ2 gene and γ2 enhancer that was missing from the clone originally sequenced.

To produce a PCR product that overlapped the previously sequenced γ3 contig, a different downstream primer, Cγ2 Exon I, was used with the Enh96 primer (Figure 3.2B). This yielded a PCR product of approximately 7.5 kb. This PCR fragment was then shotgun cloned and sequenced. After obtaining 115 sequences a 7.4 kb contig was assembled. 89 sequences were used to construct the 7.4 kb contig for an average redundancy of 7.8. 20 sequences were of M13 sequencing vector and 6 sequences did not fall into any significant contig. The sequence of the PCR product overlapped the first exon of Cγ2 with 100% similarity and the relative distances of the Cγ2 exons and enhancer were found to match those values previously reported (Vernooij *et al.*, 1993; Spencer *et al.*, 1991).

It appears that the γ3 cosmid may be unstable and prone to rearrangement. The contamination of the original clone stocks may even represent clones that have lost or modified the cosmid insert due to rearrangement, but still maintain antibiotic resistance and can therefore outgrow those clones which still carry the unchanged cosmid.

3.1.2 Gap Filling by PCR, Sequencing and Assembly

PCR was also used to fill the gap between the γ3 and γ84 cosmids. Primers were constructed from the 3' end of the γ3 contig and the 5' end of the γ84 contig. Using a long PCR protocol, a band of approximately 4.5 kb was obtained (Figure 3.4). A second band of weak intensity was observed in the sample PCR, but was not present in the negative control. The PCR product was shotgun sequenced and a 4.6 kb contig was assembled after obtaining 98 sequences. 77 sequences were used to construct the 4.6 kb contig for an average redundancy of 9.3. 14 sequences were of M13 sequencing vector and 7 more sequences were not entered in any significant contig.

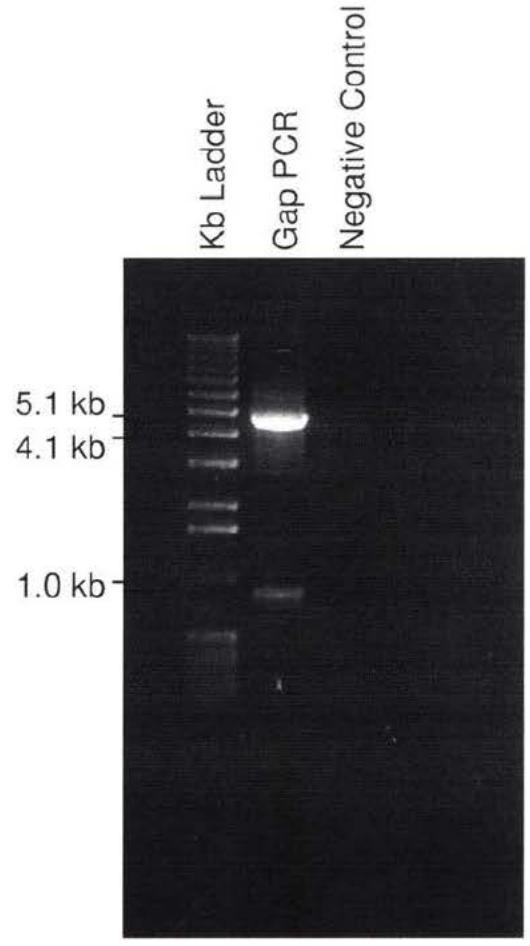

Figure 3.4. PCR amplification of a 4.5 kb fragment of cosmid γ74. Primers were constructed from the facing ends of cosmid γ3 and cosmid γ84. A lower intensity band below 1 kb in size may have been a product of non-specific primer binding.

3.1.3 Verification of Sequence

Following sequencing and assembly of the cosmids and PCR products, all of the resulting contigs were aligned to form a single contiguous sequence 62,031 bp in length. To determine if this sequence accurately represented the genomic DNA, it was tested by comparison with the physical map, restriction enzyme analysis, comparison of PCR product lengths and direct sequence comparison.

The experimental sequence was first aligned with TCRγ cDNA sequences obtained from Genbank. Dot matrix alignments indicated the position of the exons in the experimental sequence and these positions were found to match the relative positions predicted in the physical map of the locus (Vernooij *et al.*, 1993). The cDNA sequences also showed 98% to 100% identity with the exons in the experimental sequence (data not shown). Previous mapping studies of the mouse TCR γ locus have published restriction map data for portions of the locus (Hayday *et al.*, 1985; Vernooij *et al.*, 1993). These reported restriction cut sites agree with those found to occur in the experimental sequence (data not shown). Other investigators have also reported genomic sequence for regions within that covered by the experimental sequence (Hayday *et al.*, 1985; Kuziel *et al.*, 1994). Alignment of these sequences with the experimental sequence indicate 98% to 100% identity (data not shown). Finally, BLAST searches of the intervening DNA consistently return alignments of high similarity with mouse genomic DNA and common mouse specific interspersed repetitive elements. These data suggest that the sequence obtained in this series of experiments is representative of the corresponding mouse genomic DNA.

3.2 Sequence Analysis

3.2.1 Gene organization and mapping

Alignment of cDNA sequences with the genomic sequence allows their relative locations to be located precisely. The sequence reported here represents the first reported

genomic sequence that has linked all of the exons and gene segments of the γ2 and γ4 clusters, although partial genomic sequences have been previously reported within this region (Hayday *et al.*, 1985; Vernooij *et al.*, 1993; Kuziel *et al.*, 1994). Having linked all the exons and gene segments with continuous sequence data it was possible to determine the relative positions of these features with great accuracy (Figure 3.5).

The γ2 cluster lies in the opposite orientation to the other three clusters of the mouse TCR γ locus (Hayday *et al.*, 1985). Vγ1.2 was found to be separated from Jγ2 by 25,284 bp, and Jγ2 was separated from Cγ2 by 3692 bp. Cγ2 was previously found to have been encoded by 3 exons (Hayday *et al.*, 1985) and this was confirmed by our results. The first exon was 330 bp long and was separated from exon 2 by 1423 bp. Exon 2 was just 29 bp and was separated form the third exon by 567 bp. The third exon of Cγ2 was 140 bp in length.

In cluster γ4, the distance between Vγ1.1 and Jγ4 was previously reported as 2.2 kb (Traunecker *et al.*, 1986). In the genomic sequence this distance was found to be just 1696 bp. The distance between Jγ4 and Cγ4 was found to be 2344 bp. Cγ4 consists of 4 exons as previously reported by Vernooij and co-workers (1993). As with Cγ2, the first exon of Cγ4 was encoded by 330 bp. It was separated from the second exon by 4646 bp. The second and third exons were 54 and 53 bp respectively and were separated by 739 bp. The fourth exon was 1825 bp from the end of the third exon and was 142 bp long. The sequence between the two V gene segments agreed exactly with the sequence obtained by Kuziel and co-workers (1994), who measured the distance between the V gene segments at 2988 bp. Physical mapping efforts had previously estimated the same distance at 4 kb (Vernooij *et al.*, 1993).

3.2.2 Comparison of germline and cDNA sequences

Alignment of cDNA sequences of TCRγ mRNA with the genomic sequence of the locus allowed the characterization of the N region diversity occurring in V-J junctions

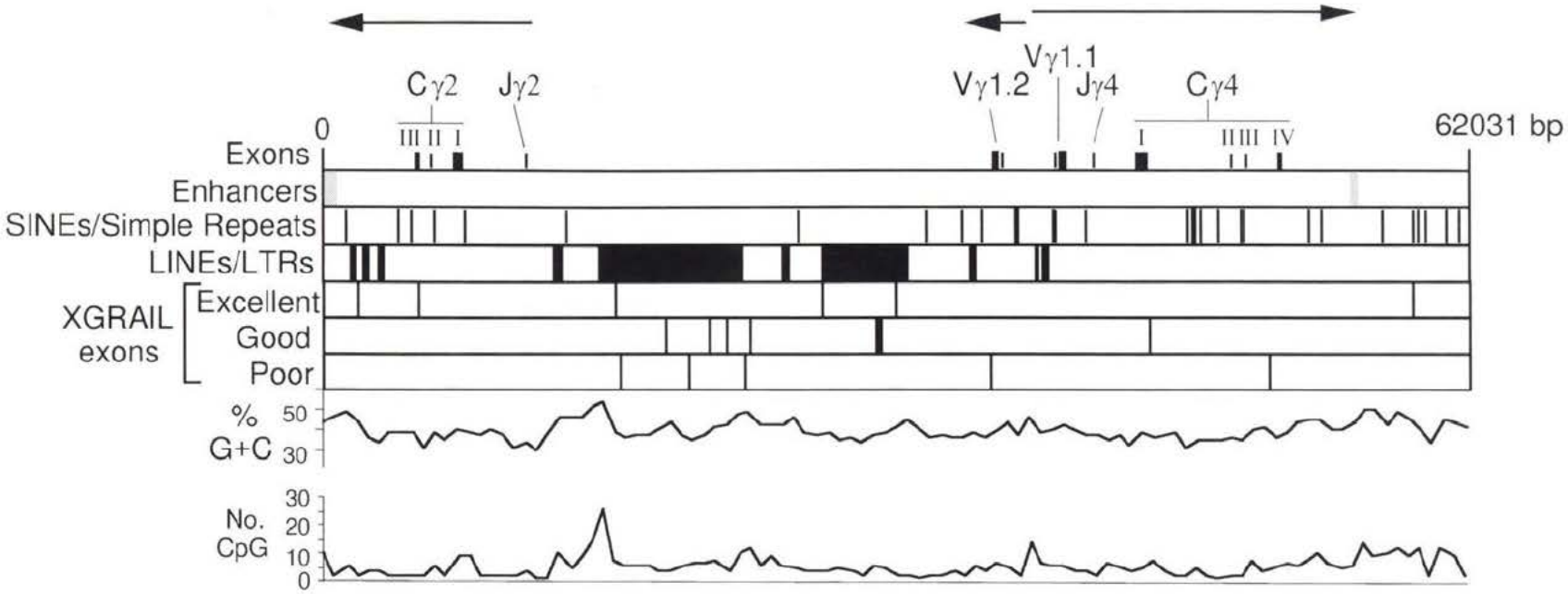

Figure 3.5. Analysis of the genomic DNA sequence spanning the mouse TCR γ2 and γ4 gene clusters. Variable (V) and joining (J) gene segments and constant region (C) genes are indicated. Exons of the C genes are numbered with Roman numerals. Arrows at the top of the figure indicate direction of transcription of the exons. SINEs, simple repeats, LINEs and LTRs are represented by black rectangles. Exons predicted by XGRAIL are separated according to the degree to which they match the optimal parameters set to define an exon. The G+C content (% G+C) and number of CpG sites (No. CpG) is shown measured over a sliding window of 800 bp with 600 bp offset.

following V(D)J recombination. cDNA sequences were obtained from Genbank using a search string of a short sequence from both the 3' end of each V and the 5' end of each J gene segment. The resulting mRNAs were then aligned with and compared to the germline V and J sequences in order to observe the extent of deletions and random nucleotide additions that have occurred at the junctions of the gene segments.

Seven Vγ1.1-Jγ4 arrangements were found in Genbank. When aligned with the germline sequences, the Vγ1.1-Jγ4 rearrangements show an average of 4.7 bp lost due to exonuclease activity at the V sequence and an average 2 bp deletion of the 5' end of the J sequence. Within the N region, 3 of 7 sequences share a TGGAT motif, and 2 share a TCGA motif (Figure 3.5A).

Twelve rearrangements of Vγ1.2 with Jγ2 were found in Genbank. These show an average 3.4 bp of exonuclease digestion at the 3' end of the V. Three of the cDNA sequences share an N region sequence of AT. One cDNA has an N region insertion of a single T while 3 contain apparently additions of 5 or 6 nucleotides. The 5 remaining cDNAs contain no N region nucleotides. At the 5' end of Jγ2 there is an average exonuclease digestion of 0.3 bp, with just 3 of the 12 sequences having undergone any exonuclease digestion at all (Figure 3.5B).

It must be considered that these data are based on a small number of samples. The 3 sequences having a TGGAT motif come from a single study of T cell diversity following radiation exposure and thus represent clonal populations which have expanded due to exposure to a common stimulus (Kishihara *et al.*, 1988).

3.2.3 Repeat element location

Genome wide repeats make up the majority of identifiable elements in the genomic DNA of vertebrates, but their function in the genome is not understood. Genome wide repeats can be classified as belonging to particular families. These families include long interspersed nuclear elements (LINEs), long terminal repeats (LTRs), and short

A

	Vγ1.1	EV	N	EJ	Jγ4
Germline	GAAGATGAAGCCACCTACTACTGTGCAGTCTGGATAAA			TCAGGCACATCATGGGTCAAGATATTTGCC	
AKR	GAAGATGAAGCCACCTACTACTGTGCAGTCTGGATAA			GCACATCATGGGTCAAGATATTTGCC	
RAG12	GAAGATGAAGCCACCTACTACTGTGCAGTCTG		TTGGATGAA	AGGCACATCATGGGTCAAGATATTTGCC	
RAG14	GAAGATGAAGCCACCTACTACTGTGCAGTCTG		TTGGATAAGG	GGCACATCATGGGTCAAGATATTTGCC	
RAG43	GAAGATGAAGCCACCTACTACTGTGCAGTCTG		TTGGAT	AGGCACATCATGGGTCAAGATATTTGCC	
MSGV1	GAAGATGAAGCCACCTACTACTGTGCAGTCTGGATA		TAG	GGCACATCATGGGTCAAGATATTTGCC	
MMTRG02	GAAGATGAAGCCACCTACTACTGTGCAGTCTG		TCGA	TCAGGCACATCATGGGTCAAGATATTTGCC	
TCGXE	GAAGATGAAGCCACCTACTACTGTGCAGTCTG		TCGA	TCAGGCACATCATGGGTCAAGATATTTGCC	

B

	Vγ1.2	EV	N	EJ	Jγ2
Germline	GAAGCCACCTACTACTGTGCAGTCTGGATGAG			ATAGCTCGGGCTTTCACAAGGTATTTGCA	
TCRG04	GAAGCCACCTACTACTGTGCAGTCTGGAT			ATAGCTCGGGCTTTCACAAGGTATTTGCA	
RAG11	GAAGCCACCTACTACTGTGCAGTCTGGAT		AT	ATAGCTCGGGCTTTCACAAGGTATTTGCA	
RAG4	GAAGCCACCTACTACTGTGCAGTCTGGAT			TAGCTCGGGCTTTCACAAGGTATTTGCA	
RAG42	GAAGCCACCTACTACTGTGCAGTCTGG		GGGAGG	AGCTCGGGCTTTCACAAGGTATTTGCA	
RAG50	GAAGCCACCTACTACTGTGCAGTCCGG		T	ATAGCTCGGGCTTTCACAAGGTATTTGCA	
TCEL05	GAAGCCACCTACTACTGTGCAGTCTGGATGAG		AT	ATAGCTCGGGCTTTCACAAGGTATTTGCA	
tcrg1p	GAAGCCACCTACTACTGTGCAGTCTGGATGAG			ATAGCTCGGGCTTTCACAAGGTATTTGCA	
TCRG4P	GAAGCCACCTACTACTGTGCAGTCTGGATGAG		CCCCG	TAGCTCGGGCTTTCACAAGGTATTTGCA	
TCRG	GAAGCCACCTACTACTGTGCAGTC			ATAGCTCGGGCTTTCACAAGGTATTTGCA	
TCGXC	GAAGCCACCTACTACTGTGCAGT		AT	ATAGCTCGGGCTTTCACAAGGTATTTGCA	
TCGXB	GAAGCCACCTACTACTGTGCAGTCTGG		CCCAT	ATAGCTCGGGCTTTCACAAGGTATTTGCA	
TCGXD	GAAGCCACCTACTACTGTGCAGTCTGGATGAG			ATAGCTCGGGCTTTCACAAGGTATTTGCA	

Figure 3.6. Diversity at V-J junctions. A. Alignment of Vγ1.1-Jγ4 cDNA sequences with germline junctions. B. Alignment of Vγ1.2-Jγ2 cDNA sequences with germline junctions. EV, region of exonuclease activity on the 3' end of the V gene segments. EJ, region of exonuclease activity on the 5' end of the J gene segments. N, region of random nucleotide addition.

interspersed nuclear elements (SINEs). These repeats are copied throughout the genome and some loci contain larger numbers of repeats than others. Identification of these repeats in large scale sequencing projects may be useful in determining patterns of repeat duplication and insertion which may aid in determining their effect on organisms and the evolution of the genome.

Repeat elements located by Repeat Masker were separated on the basis of size. LINEs and LTRs are shown on one line while SINEs and simple repeats are shown on another (Figure 3.5). Just downstream of Jγ2 lies an LTR insertion sequence which is followed closely by a simple GA repeat. About 1.8 kb downstream lies a 7.6 kb LINE of the type L1. Another LTR insertion sequence lies 2.3 kb downstream and is followed closely by a simple TC repeat. A shorter LINE of 4.8 kb, also an L1, follows 1.2 kb further downstream. The SINE and simple repeats seem to be distributed relatively evenly throughout the sequence. Although undetected by Repeat Masker, an imperfect tandem repeat of $(TACTGCCCTTCAGGACTCAAT)_{14}$ was observable on a dot matrix plot located between the two LINE sequences.

Genome-wide repeats, such as SINEs, LINEs, and LTRs were found to represent 25% of the total sequence (Table 3.1). Due largely to their size, the largest contributors were LINEs, which by themselves represent 20.4% of the total sequence. LTRs and retroviral sequences represented 1.8% of the sequence and SINEs accounted for 1.3%. It is interesting that two complete L1 repeats were found between the Jγ2 and Vγ1.2 gene segments and it appears that no coding regions have been interrupted by the insertion of repetitive elements.

3.2.4 Analysis of internal homology

The evolution of the genome proceeds by gene duplications, deletions, rearrangements, gene conversion and spontaneous mutation. Regions of the genome that

Table 3.1. Distribution of genome-wide repeat elements. Genome wide repeats were identified by screening the TCR γ sequence against a database of interspersed repeat consensus sequences using Repeat Masker.

Type	Name	Copies	Proportion of sequence (%)
LINEs			
	MML	4	20.4%
	L1	3	1.2%
LTR			
	MMLTRIS	4	1.4%
	Retroviral	1	0.4%
SINEs			
	MMFLAM	1	0.1%
	MMB1	3	0.5%
	MMB2	2	0.4%
	MIR	2	0.3%
Unclassified			
	SQR	2	0.3%
Total		22	25.0%

have a critical function remain highly conserved in sequence, whereas those regions that are not critical to the survival of the organism accumulate mutations and rapidly diverge in sequence homology. Comparisons of genomic sequences provide clues to the evolutionary events occurring within the genome. Similar islands of non-coding DNA indicate recent duplications and perhaps sequences with some function. In the 3' end of the mouse TCRγ locus, a dot matrix analysis comparing the genomic sequence with itself was used to determine the overall similarity of the γ2 and γ4 gene clusters, locate conserved regulatory elements, flanking regions and other unknown conserved duplication products.

Little or no homology was noted between the Cγ2 and Cγ4 genes outside of the exons themselves (Figure 3.7). Some very low homology was noted in the 3' flanking region that extended to include the γ2 enhancer. To investigate this further, a dot matrix plot of the genomic sequence with the γ1 enhancer (γ1E) sequence (Spencer *et al.*, 1991) was performed. Strong homology with the putative γ2 enhancer (γ2E) (Vernooij *et al.*, 1993) was observed as expected but a second line of homology was also detected downstream of Cγ4 (Figure 3.8). This suggested the presence of an as yet undescribed putative γ4 enhancer element. While the other three enhancers found to date show >96% homology to the γ1 enhancer (Vernooij *et al* ., 1993), the putative γ4 enhancer shares just 65% homology over a 340 bp region (Figure 3.9). This region was found to lie approximately 3.5 kb downstream of the stop codon of Cγ4. The weak homology between the introns and flanking regions of Cγ2 and Cγ4 suggests that they are the products of a very old duplication event.

In contrast to the C genes, the two V gene segments, Vγ1.1 and Vγ1.2 showed very high similarity both inside and outside of the coding regions (Figure 3.10). The V regions appear as two closely opposed inverted duplication products. An area of 91% homology extended approximately 1000 bp upstream from the 5' untranslated region through the coding regions and over 300 bp 3' of the recombination signals (Figure 3.11).

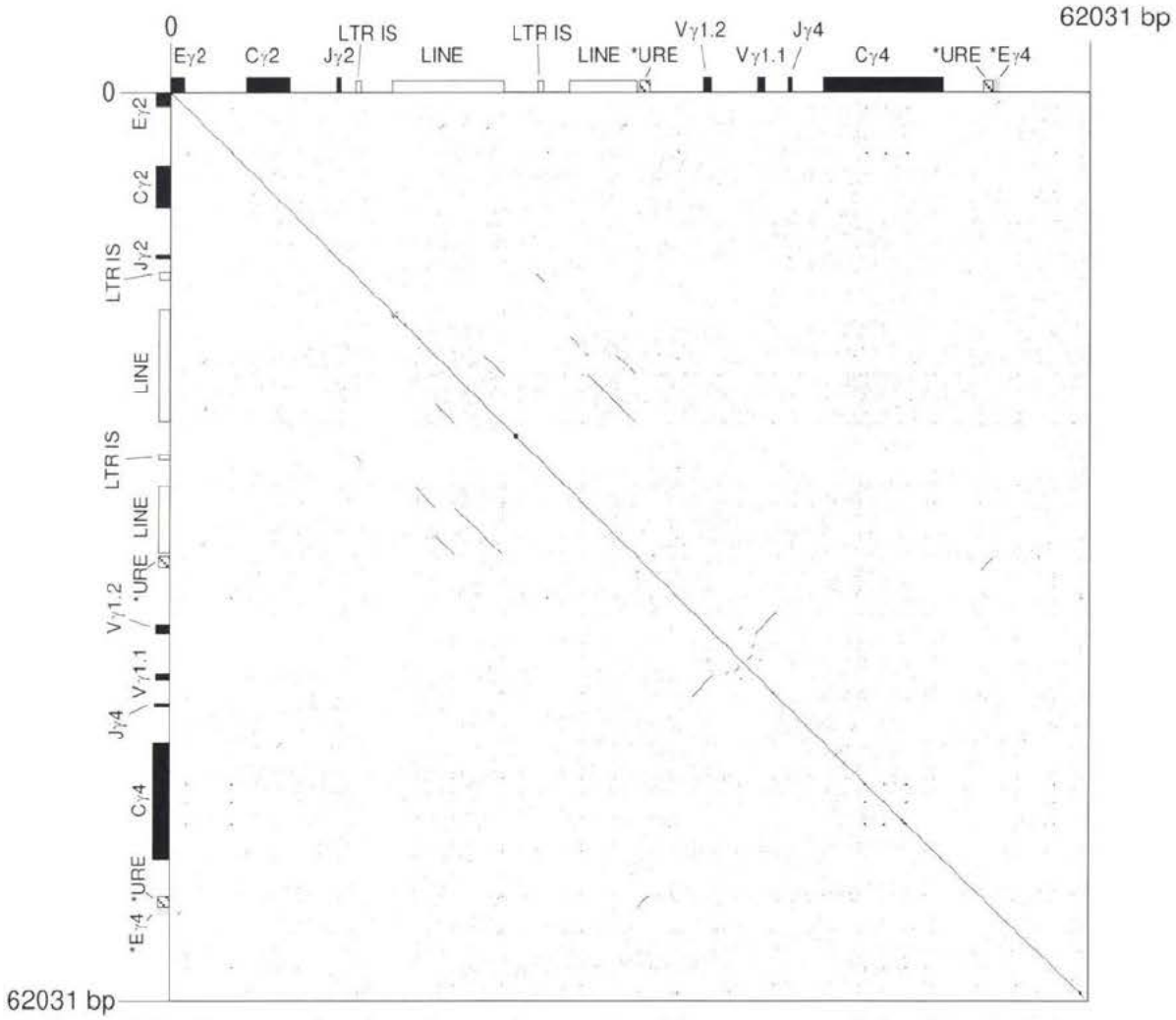

Figure 3.7. Dot matrix plot of complete nucleotide sequence of the 3' end of the mouse TCR γ locus against both the coding and template strands of the same region. Homologous regions occurring in tandem form diagonal lines from top left to lower right, while homologous regions occurring in an inverted orientation appear as diagonal lines from top right to lower left. The locations of the genes, including introns, are shown along the top and left hand sides in black while the locations of LINE and LTR repeats are shown as hollow boxes. The putative γ4 enhancer is shown as a shaded box and an inverted repeat of unknown origin is shown as a hatched box. *, putative features. URE, Unclassified Repeat Element.

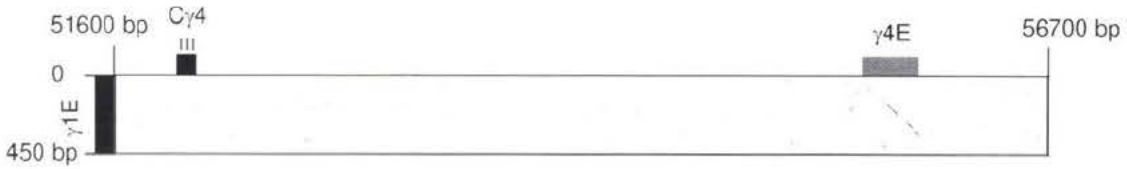

Figure 3.8. Dot matrix plot of the 3' flanking region of the γ4 gene cluster against the γ1 enhancer sequence. The putative γ4 enhancer is shown as a shaded box. Black dots indicate 100% homology over a 25 bp window, lower homology scores are depicted as progressively lighter shades of grey. Diagonal lines from top left to lower right indicate homology between coding strands. Diagonals from top right to lower left indicate homology between coding and non-coding strands.

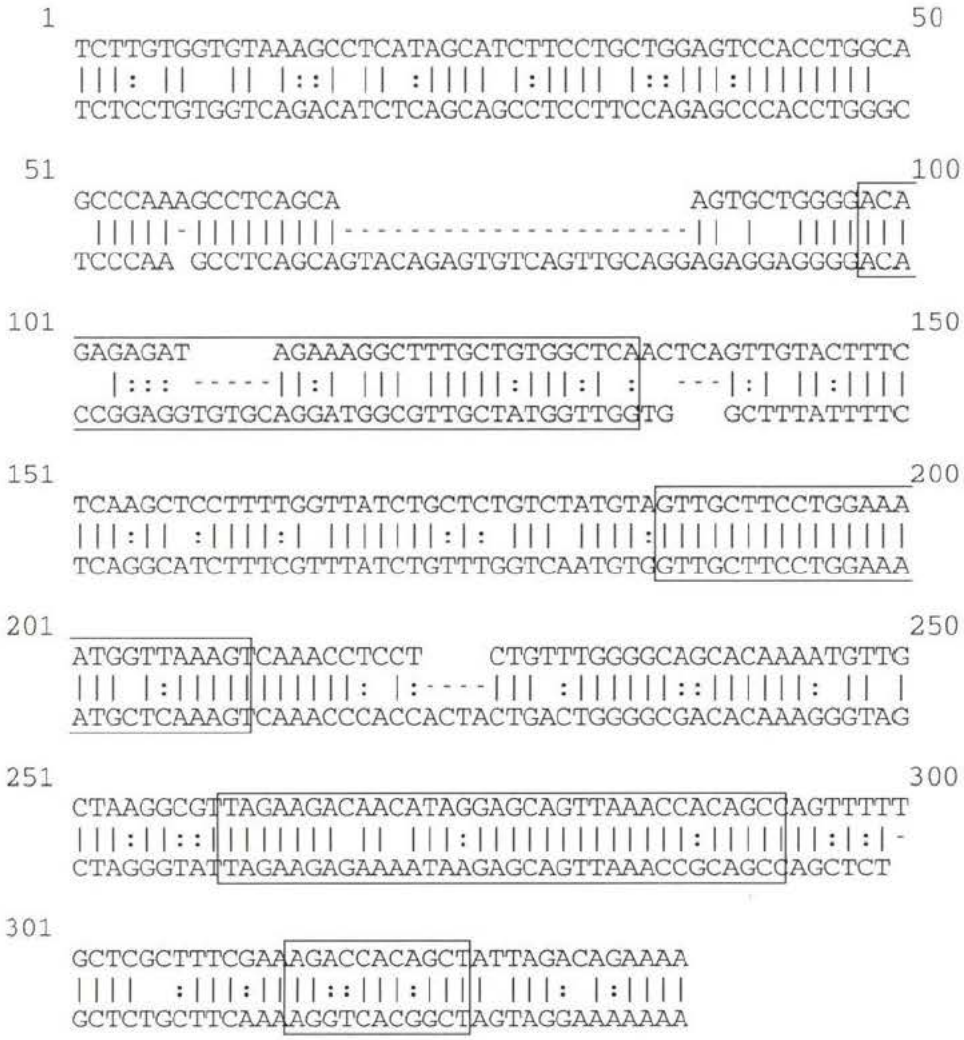

Figure 3.9. Alignment of the mouse γ1 enhancer with the putative γ4 enhancer. The top sequence is the γ1 enhancer and the bottom is the γ4 enhancer. Identical matches are indicated by a vertical line. Differences arising by transition are indicated by two dots and differences arising by transversion are indicated with a blank space. Insertions or deletions are indicated by dashes. Boxed sequences are those which may be important for enhancer function (Spencer *et al.*, 1991).

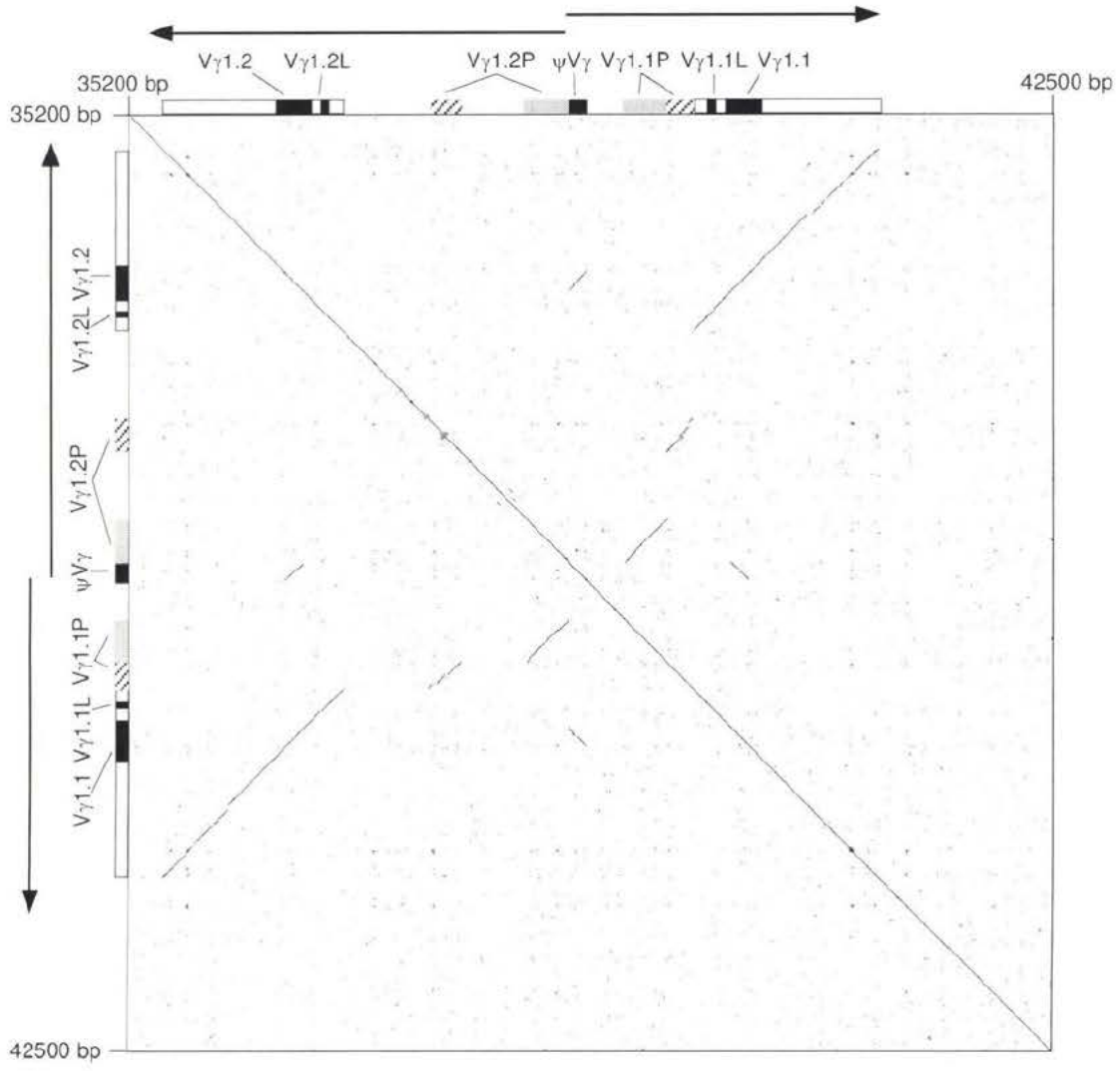

Figure 3.10. Dot matrix self comparison of the inverted homologous regions covering the Vγ1.1 and Vγ1.2 genes segments. Exons are shown as black boxes at the top and left side of the plot. Non coding sequence immediately surrounding the exons is shown in bordered light grey, while two distinct blocks of homology 5' of the start codons of Vγ1.1 and Vγ1.2 are indicated in dark grey and a cross hatched pattern. Homologous regions occurring in tandem form diagonal lines from top left to lower right, while homologous regions occurring in an inverted orientation appear as diagonal lines from top right to lower left. Arrows indicate the direction of transcription. P denotes a promoter sequence, L denotes the leader sequence, and ψ indicates a psuedogene.

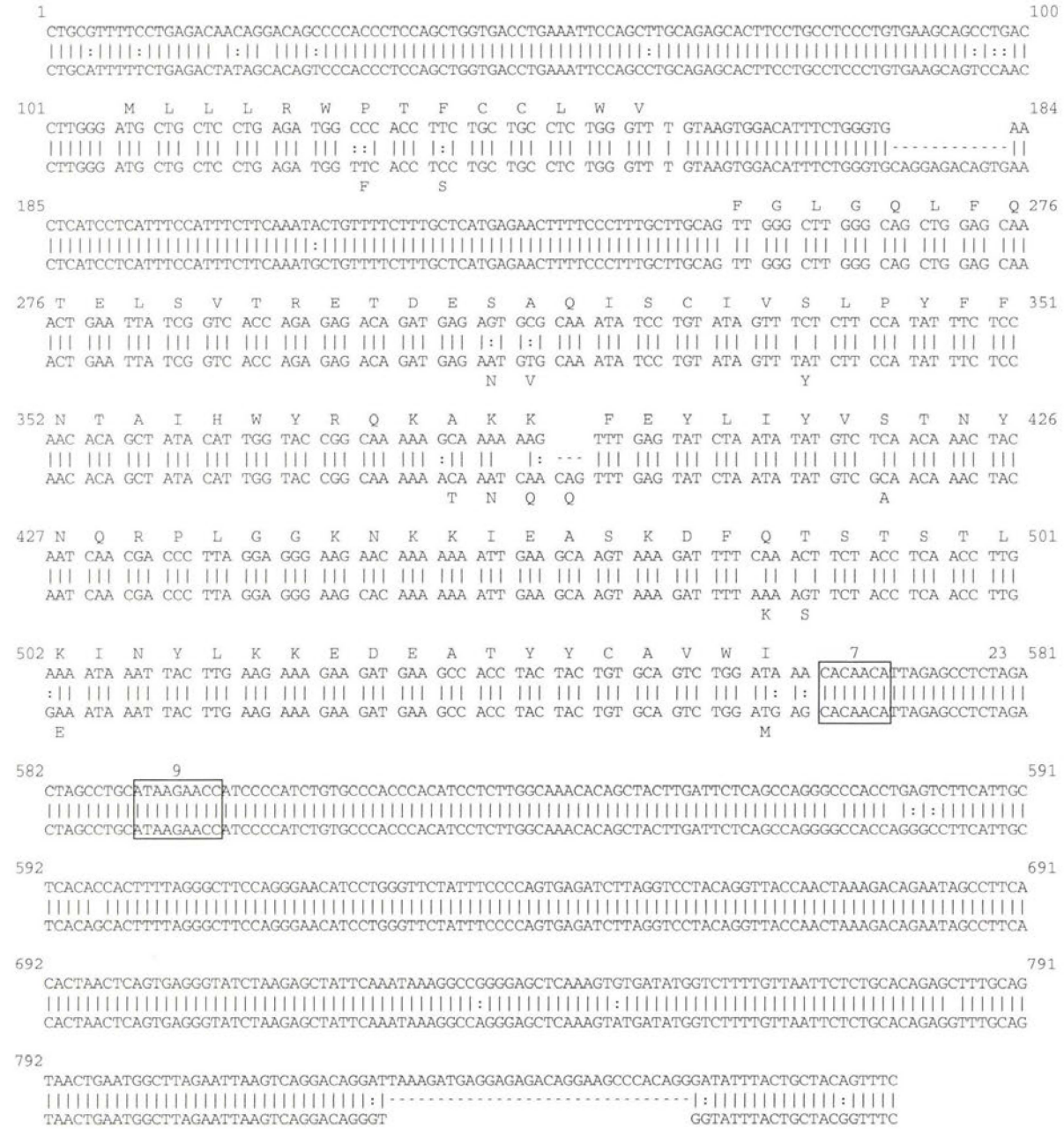

Figure 3.11. Alignment of the homologous sequences surrounding Vγ1.1 and Vγ1.2. The top sequence is Vγ1.1 and the bottom sequence is Vγ1.2. Identical matches are indicated by a vertical line. Differences arising by transition are indicated by two dots and differences arising by transversion are indicated with a blank space. Insertions or deletions are indicated by dashes. Codons are separated by spaces and the translation shown above each codon. Where the two amino acid sequences differ, the difference is shown beneath the lower sequence. Boxed sequences are heptamer and nonamer recombination signals separated by 23 bp.

Within the intron following the first exon, or leader sequence, a 12 base insertion or deletion occurred and within the conserved sequence 3' of the coding region, a 31 base insertion or deletion was present. The sequence obtained for the 5' end of the V gene segments and the intervening DNA between Vγ1.1 and Vγ1.2 was identical to the sequence obtained previously by Kuziel and co-workers (1994). The very high similarity of the V genes suggests that unlike Cγ2 and Cγ4, the V genes are the product of very recent duplication and inversion events.

The most prominent of the duplications observed in the dot matrix plot occurred within and among the large LINE repeats located between Jγ2 and Vγ1.2 (Figure 3.12). The first duplicated elements are the LTR insertion sequences which precede each of the LINE repeats. A small tandem duplication of about 250 bp was observed at the 5' end of the first LINE. The two LINEs shared substantial homology which was broken up by the duplication of about 1000 bp from the 3' end of the first LINE and reinserted into its middle. Over 1000 bp at the 5' end of the first LINE had no homologue in the second LINE.

Outside of the exons of the C genes, the V gene duplication and LINE repeats, little internal homology was noted. One exception was an inverted repeat of approximately 800 bp, one of which was found to border 5' side of the putative enhancer of Cγ4 and the other bordered the 3' end of the downstream of the two LINE repeats (Figure 3.7). The two sequences shared 89% sequence identity (Figure 3.13) . A BLAST search of Genbank with this sequence produced no probable matches. A heptamer, AGAAAAA, was found at the 5' end and an inverted homologue, TTTTTGT, was found at the 3' end. At this time the origin of this sequence is not known.

3.2.5 Prediction of exons

In a large genomic sequence there is a chance of finding an unmapped or undiscovered gene. To ensure that all genes in the sequence are located it is necessary to

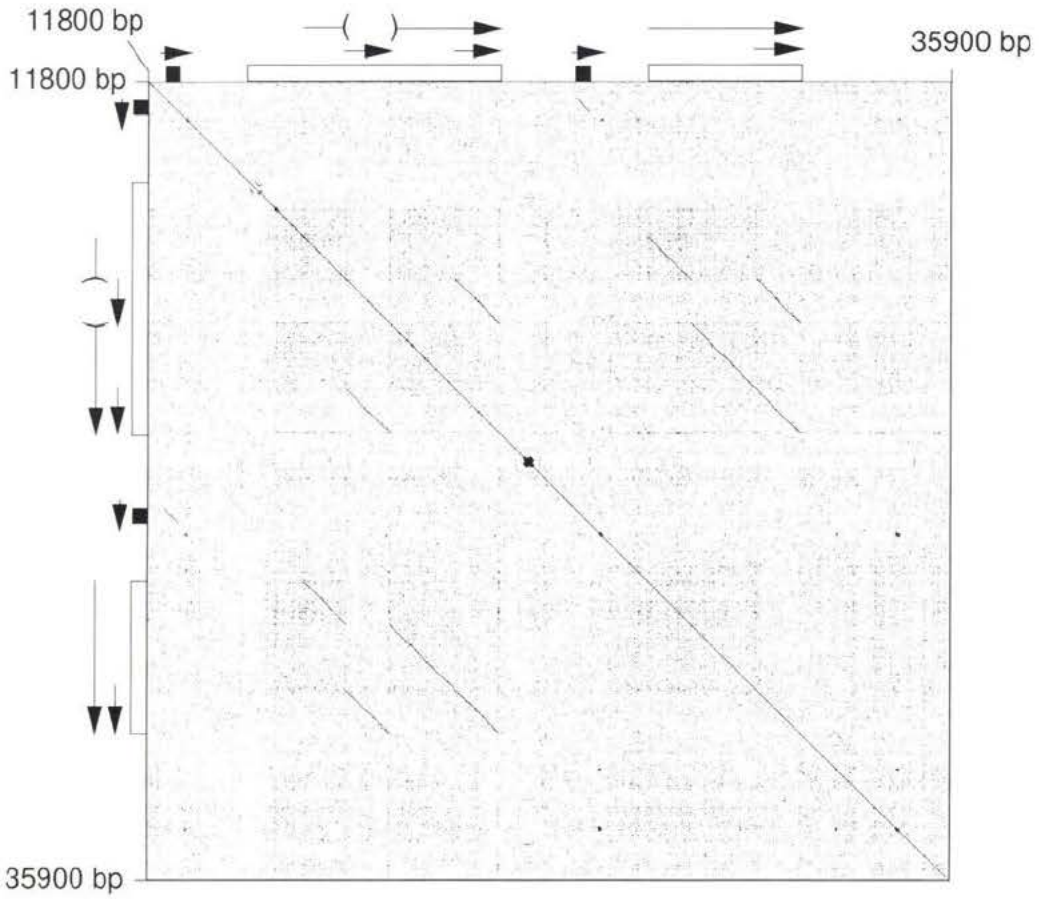

Figure 3.12. Dot matrix self comparison of the region between the Jγ2 and Vγ1.2 gene segments containing LTR and LINE repeats. LTRs are indicated by black boxes and LINEs are indicated by empty boxes. Homologous regions shared between the two LINEs are indicated by arrows located above the LINEs at the top and the left hand side of the plot. Homologous regions occurring in tandem form diagonal lines from top left to lower right, while homologous regions occurring in an inverted orientation appear as diagonal lines from top right to lower left. Brackets indicate an insertion of a repeated unit into a larger repeat.

Figure 3.13. Alignment of unidentified repeat elements (UREs). The top sequence is the first repeat element, located 3' of the LINE repeats. The bottom sequence is reverse complement of the second repeat, located at the 5' side of the putative γ4 enhancer. Differences arising by transition are indicated by two dots and differences arising by transversion are indicated with a blank space. Insertions or deletions are indicated by dashes. Heptamer poly A and poly T repeats at each end of the sequences are boxed.

search the entire sequence against known cDNA and protein sequences. This may indicate the location of an unmapped gene or a homologous gene that was previously unmapped to the genomic DNA. Additionally there may be undiscovered exons which are not present in any database. A recent approach to this problem was the development of a complex pattern recognition algorithm to recognize common features of exons. The GRAIL program at Oak Ridge National Laboratories allows free access to such an algorithm through the internet using the world wide web, electronic mail or a graphical front end called XGRAIL.

XGRAIL 1.2 analysis of the sequence predicted several exon candidates with probabilities ranging from 20% to 90%. Only one predicted exon matched any of the 13 known exons (Figure 3.5). A BLAST search of GenBank and protein databases with the other XGRAIL exon candidates produced no probable matches with any known cDNA or protein sequences (data not shown). It is possible that the pattern recognition algorithm of GRAIL was confused by the absence of inward facing splice sites on the V and J segments. To test this hypothesis the intervening sequence was removed, bringing the V and J segments together to form a single exon bordered by the appropriate RNA splice sites. This sequence was then submitted using XGRAIL, but the exon prediction was not improved (data not shown). As has been previously noted, gene finding algorithms are notably poor in finding TCR genes (Rowen *et al.*, 1996).

GC content is thought to be a useful predictor of the gene richness of a genomic locus and GC content is one of the characteristics used by XGRAIL to predict exons. The average GC content for the region is 40.7% with a standard deviation of 4.8% (Figure 3.5). Peaks in GC content and number of CpG sites do not appear to correspond with exons or gene segments. Both the largest peaks and lowest values in GC content and number of CpG sites occur in non-coding regions. Low GC content in the coding regions may be a problem for the GRAIL algorithm in this region.

To locate any known genes or their homologues that may occur within the experimental sequence, all repeat elements were removed with Repeat Masker and the sequence was broken into segments and sequentially submitted to the NCBI BLAST search of Genbank. This turned up no probable hits other than the TCR gene segments and exons already mapped to the region. BLAST/BEAUTY is a similar search system, but instead of searching the nucleotide database it translates the nucleotide sequence to amino acid sequences in all six reading frames and uses the amino acid sequences to search available protein databases. Like the BLAST search, BLAST/BEAUTY also failed to find any unknown coding regions (data not shown).

3.3 Evolutionary Analysis of the Cγ Clusters

To put the relationship between Cγ2 and Cγ4 into a broader evolutionary context, the phylogenetic relation of C gene sequences from several species was examined. This was accomplished by aligning the amino acid sequences of the C genes from various mammalian species and using the amino acid alignment as a basis for a parsimony analysis. Amino acid sequences for the Cγ genes of mouse, rat, human, pig, bovine, sheep, rabbit and minke whale were aligned (Figure 3.14). Strong similarity was noted in the extracellular and transmembrane domains of all Cγ genes, while the hinge domain was divergent even between Cγ genes of individual species.

A parsimony tree was produced from the amino acid alignment using human Cα, Cδ and Cβ as outgroups (Figure 3.15). Due to their high homology (>96% nucleotide similarity) the mouse Cγ1 and Cγ2 genes grouped together with a bootstrap value of 100%. Orthologous mouse and rat C genes grouped together with high bootstrap values (99%), but the relation between the rodent Cγ4 genes and Cγ1-like genes was not resolved. Likewise, a relationship between the human, rodent and lagomorph Cγ genes was not resolved.

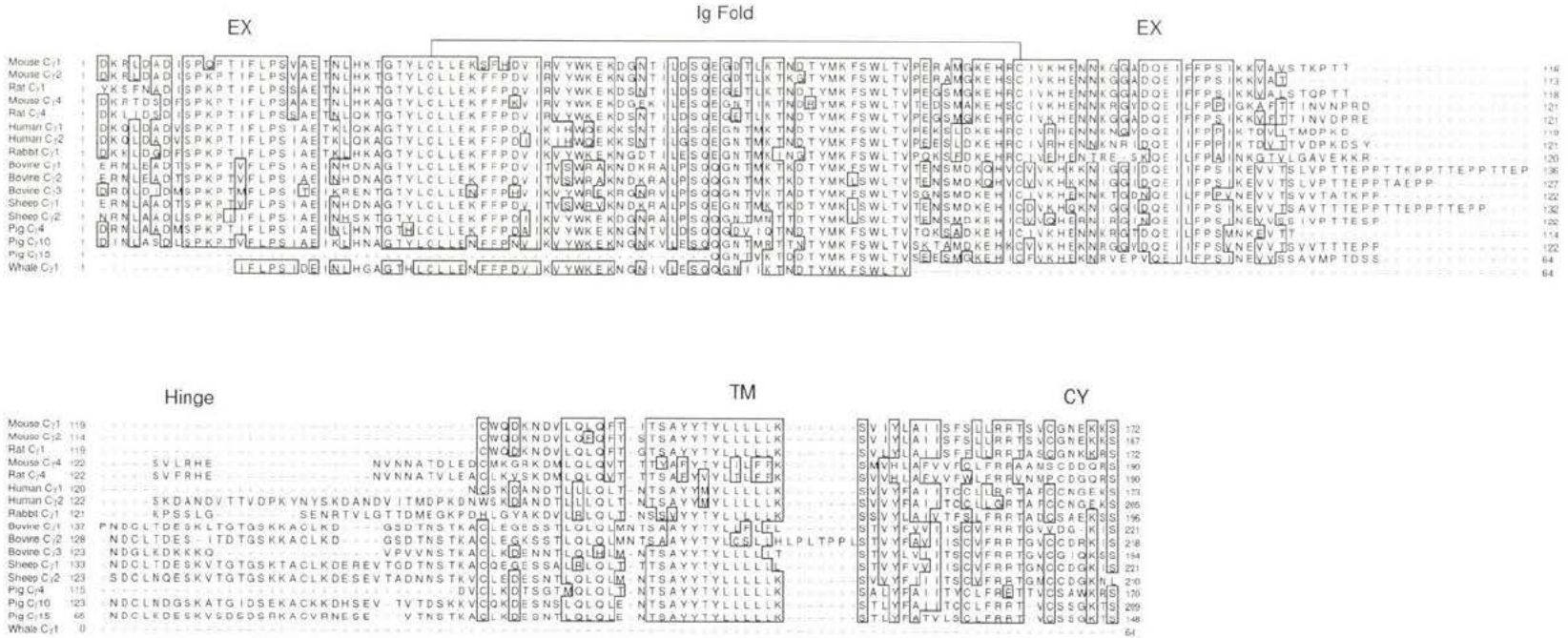

Figure 3.14. Amino acid alignment of Cγ genes from mouse, rat, human, rabbit, sheep, bovine, pig and whale. Boxes indicate identical residues. Numbers on the right and left sides show protein sequence length. Conserved cysteine residues stabilizing the Ig fold are joined by a solid line. The predicted extracellular (EX), immunoglobulin (Ig Fold), transmembrane (TM), hinge and cytoplasmic (CY) domains are indicated above the alignment.

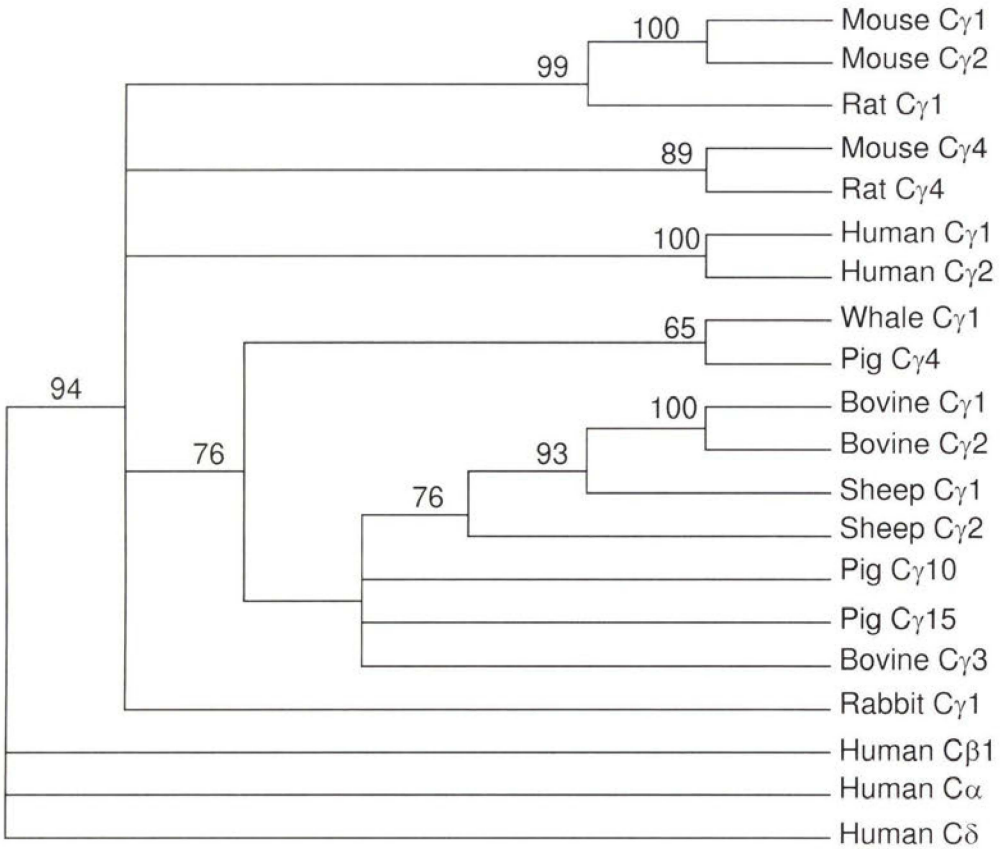

Figure 3.15. Molecular phylogeny of Cγ genes from mouse, rat, human, rabbit, sheep, bovine, pig and whale. The tree was constructed by parsimony analysis of the amino acid sequence alignment. Bootstrap values are indicated next to branch points.

The Cγ genes of the artiodactyls grouped significantly with a bootstrap of 76% (Figure 3.15). Within the artiodactyls, whale Cγ1 and Pig Cγ4 formed a sister group with a bootstrap of 65%. Bovine Cγ1 and Cγ2 grouped together with a bootstrap of 100% suggesting that they are the products of a recent duplication event. Sheep Cγ1 groups with the bovine Cγ1 and Cγ2 genes with a bootstrap of 93%. Sheep Cγ2 then joins the aforementioned clade with a bootstrap of 76%. This data suggests that the sheep Cγ1 and Cγ2 genes were duplicated before cow and sheep split from a common ancestor. The original bovine Cγ2 gene was probably deleted and the current bovine Cγ2 resulted from a recent duplication of Cγ1. The relation of the bovine Cγ3, porcine Cγ10 and porcine Cγ15 genes was unresolved.

Due to the lack of resolution between the rodent Cγ1 and Cγ4 lineages and the human Cγ1 lineages it was not possible to determine the relative time of divergence of the Cγ1-like genes and murine Cγ4. The lack of resolution suggests an ancient duplication, but whether it occurred prior to the divergence of the rodent and primate lineages or after is not clear. The nearly complete lack of similarity between the introns and flanking regions of the Cγ2 and Cγ4 genes observed in the genomic sequence supports this conclusion.

DISCUSSION

We have reported the determination of 62 kb of contiguous genomic DNA sequence from the mouse TCR γ locus. This sequence, extending from the γ2 enhancer to the 3' end of the locus, represents the third largest mouse genomic sequence reported to date and confirms the organization of the locus as previously reported by physical mapping (Vernooij *et al.*, 1993; Raulet, 1989; Traunecker *et al.*, 1986; Iwamoto *et al.*, 1986; Hayday *et al.*, 1985). The genomic DNA sequence provides a fine scale map with which to investigate the organization of this locus and its evolutionary history.

4.1 Sequence Determination

The contig assembly for the γ84 cosmid gave a contiguous sequence of the expected length, but the sequence of the γ3 cosmid was found to be incomplete. Instability of cosmids overlapping the 5' end of the γ3 cosmid was previously noted (Vernooij *et al.*, 1993), but the fact that cosmid γ3 had rearranged went undetected until it had been completely sequenced. Restriction digests of the γ3 cosmid showed a clear banding pattern that added up to the expected size of the insert, but attempts to construct a restriction map with digests of neighbouring cosmids were inconclusive. BLAST searches with sequenced M13 subclones of γ3 matched expected genes and mouse genomic DNA, suggesting that the γ3 clone was correct. After the sequences were assembled, however, it was discovered that more than half of the expected insert was missing and that a *Tn*10 IS element was inserted between the first exon of Cγ2 and the interior of the cosmid vector. Although a portion of the missing region spanning from the first exon of Cγ2 to the end of the γ2 enhancer was obtained by PCR, up to 15 kb of the γ3 cosmid remains missing from this analysis. Sequencing beyond the middle of the γ3 will require a modified sequencing strategy which circumvents the problem of cosmid stability.

Aside from the loss of the insert 5' of the first exon of Cγ2, the sequence obtained from γ3 appeared unaffected by the rearrangement. The 3' end of the sequence overlapped

with cosmid γ74 as expected, enabling the PCR amplification of the gap between γ3 and γ84. The 5' end matched the previously determined genomic sequence for Cγ2 determined by Hayday and co-workers (1985) and the sequence spanning the Vγ1.2 and Vγ1.1 gene segments agreed with the sequence obtained by Kuziel and co-workers (1994). Restriction enzyme cut sites and relative locations of the gene segments and exons within the genomic sequence also agreed with those determined previously (Vernooij *et al.*, 1993).

The mechanism by which the rearrangement of the γ3 clone occurred is not known, however it is interesting that no stable cosmids were available for the region 5' of the γ2 cluster and 3' of the γ3 cluster. This suggests that the sequence in this region is inherently unstable in *E. coli*. It may also contribute to instability in the mouse TCRγ locus.

4.2 Organization and Feature Identification

Having obtained the genomic DNA sequence spanning the γ2 and γ4 gene clusters, it was necessary to locate the coding regions and regulatory sequences. All of the previously described features of the region were located by aligning cDNA and genomic sequences obtained from GenBank. These features included the Vγ1.2 gene segment, the Jγ2 gene segment, the Cγ2 gene and a putative γ2 enhancer and, at the 3' end, Vγ1.1, Jγ4 and Cγ4 (Vernooij *et al.*, 1993; Raulet, 1989; Garman *et al.*, 1986; Traunecker *et al.*, 1986; Iwamoto *et al.*, 1986; Hayday *et al.*, 1985). It was previously suggested that other rarely arranging and/or non-functional V gene segments might occur in this region (Vernooij *et al.*, 1993; Lefranc *et al.*, 1986; Forster *et al.*, 1987; Huck *et al.*, 1987; Chen *et al.*, 1988). To investigate this possibility, the region was searched using dot-matrix analysis for internal homology, BLAST searches for homologous sequences in GenBank, and BLAST/BEAUTY translation searches for homologous protein sequences. Aside from a 148 bp Vγ1 pseudogene previously described by Kuziel and co-workers (1994), no V genes other than Vγ1.1 and Vγ1.2 were found in this study.

In the physical map of the mouse TCR γ locus, the three highly homologous clusters, γ1, γ2 and γ3, were found to contain homologous enhancer sequences following their C genes, but no enhancer was mapped to the γ4 cluster (Vernooij *et al.*, 1993). Using the genomic sequence determined in this study, a putative enhancer was located downstream of the poly-A signal of the Cγ4 gene, having 65% homology to the γ1 enhancer (γ1E) (Spencer *et al.*, 1991). The level of conservation of this region in contrast to the lack of homology between the introns of Cγ2 and Cγ4 and the fact that all of the suspected functional sites identified by Spencer and co-workers (1991) fall within this 322 bp region of homology (Figure 3.9) suggest that this putative enhancer may be functional. Studies to assess the function of this putative enhancer are currently underway.

The genomic DNA sequence data obtained in this study provided a means by which the N region diversity of available mouse TCR γ cDNAs could be examined. When cDNAs containing Vγ1.1-Jγ4 rearrangements were recovered from GenBank, it was found that five of seven sequences shared a 6 bp exonuclease deletion from the 3' end of the V gene segment. Three of those five sequences shared an insertion of TTGGAT in the N region and the other two shared an N region insertion of TCGA (Figure 3.6A). Neither of these shared N region sequences was of genomic origin. The similarities in the N regions of these cDNAs probably reflect the fact that the number of cDNAs available was low and that those sequences sharing similar N region insertions were obtained from the same laboratory in response to controlled antigen exposures. Since identical V and J gene segments were used in all these clones, the similar N region motifs may reflect common binding specificities. Vγ1.2-Jγ2 rearrangements showed less exonuclease digestion of the J gene segment and less nucleotide addition on average.

Dot matrix analyses were used to analyze internal homology within the genomic DNA sequence. An important result of these analyses was the lack of conservation observed between the intervening DNA sequences surrounding the exons of Cγ2 and Cγ4

(Figure 3.7). Only the exons were similar between the two C genes. The lack of similarity outside the coding regions of the two constant region genes indicated that the two genes had diverged from an ancient duplication event.

In contrast to the C genes, the two V genes, Vγ1.1 and Vγ1.2, shared 91% homology over approximately 970 bp covering the exons, introns and 5' and 3' non-coding regions (Figure 3.10). On a dot matrix plot, the two V genes appear as a strongly conserved and undisrupted end-to-end inverted repeat. As previously reported, the genomic DNA between the V genes was found to contain two regions of homology associated with the 5' promoter region of each V gene (Kuziel *et al.*, 1994). These regions were contiguous at the 5' end of Vγ1.1 but at the 5' end of Vγ1.2, they were separated from each other by a 522 bp insertion and from Vγ1.2 by about 700 bp. A fragment of a Vγ coding region was found between the promoters of Vγ1.2 and Vγ1.1 in the same orientation as Vγ1.1 (Kuziel *et al.*, 1994). The strong homology between the V gene segments suggest a recent duplication and inversion event but evidence suggesting a mechanism by which this duplication occurred was not apparent.

To search for undiscovered coding regions, the GRAIL exon prediction system was used. GRAIL was unsuccessful at predicting exons known to exist in the sequence and those exons it predicted produced no significant matches with sequences in GenBank. The failure of GRAIL to predict exons in this locus may reflect the unusual nature of the TCR genes. The GRAIL algorithm is designed to recognize open reading frames in a variable length window and looks for patterns of open reading frames bounded by a start codon, appropriate splice sites and a stop codon. The TCR gene segments are unusual in that they do not follow this pattern. The second exon of each V gene segment ends in a recombination signal rather than a donor splice site and each J gene segment begins with a recombination signal rather than an acceptor splice site (Tonegawa, 1983). It was thought that this unusual pattern of splice sites might cause the GRAIL algorithm to overlook these

genes. This was tested by removing the intervening sequence between the V and J gene segments, creating a normal exon-intron structure with appropriate splice sites. This did not improve the exon predictions of GRAIL. Therefore, it appears that the lack of appropriate splice sites is not solely responsible for the failure of the GRAIL algorithm to predict the exons of the TCR γ genes. Other unusual features of the TCR γ locus that may contribute this problem are low GC content of the region (40.7%), and the TATA-less promoters (Kuziel *et al.*, 1994).

Another problem was the fact that GRAIL predicted 18 probable exons within the non-coding DNA which did not show homology to any known exon sequences. That 12 of the 18 probable exons occurred near or within long interspersed nuclear elements (LINEs) suggests that a high frequency of LINEs or retroviral elements in the region may cause GRAIL to return spurious results. This can be avoided by screening repeats prior to submitting a sequence to GRAIL.

4.3 Repetitive Elements

Two LINEs were found to have integrated into the intervening sequence between Jγ2 and Vγ1.2, accounting for the unusually large V-J distance in the γ2 cluster. One LINE, 7.6 kb in length, was followed closely by another LINE, 4.8 kb in length. Together these two repeat elements account for approximately 20% of the total sequence reported. The larger of the two LINEs contained 4.8 kb of sequence at its 3' end that was highly homologous to the smaller LINE with the exception of the insertion of a 250 kb block duplicated from its 3' end. The high homology of the two lines and the internal repeats within the larger LINE suggested that these elements have undergone several recent duplications and may therefore add to the instability of the locus.

The presence of LINE repeats in this region is not unusual. LINE elements were found to represent 13% of 685 kb of the human TCR β locus (Rowen *et al.*, 1996). The presence of these long repeat regions may facilitate evolutionary change through

homologous but unequal crossing over or by catalyzing translocations or rearrangements (Rowen *et al.*, 1996). Fitch *et al.* (1991) proposed a mechanism by which the γ-globin gene was duplicated through an unequal crossing over event mediated by LINE1 repeats in an ancestor of the simian primates. The presence of numerous LINE repeats in the TCRγ locus may therefore contribute to the instability of the region.

In addition to repeats belonging to previously characterized families, a novel repeat element was discovered. This element occurred as an inverted repeat of approximately 900 bp with one copy located close to the 3' end of the LINE repeats and another at the 3' end of the Cγ4 gene. The repeat sequence lacked homology with any sequence contained in GenBank and it is not known if it occurs outside of the mouse TCR γ locus.

4.4 Evolution of the TCR γ region

The lack of homology between the introns and flanking DNA of the Jγ2-Cγ2 and Jγ4-Cγ4 regions suggested that the two C genes and J gene segments were the products of an ancient duplication event. To investigate the relationship of the Cγ2 and Cγ4 genes in a broader evolutionary context, the coding regions of C genes from eight mammalian species representing four orders were compared.

Alignment of the protein sequences of Cγ genes were used to construct a parsimony tree. Mouse Cγ1 and Cγ2, which are 96% similar at the nucleotide level, grouped closely together with each other and the Cγ1 gene of the rat. Mouse Cγ4 grouped with rat Cγ4, but a point of divergence of the Cγ4 genes from the Cγ1 genes was not resolved. Human Cγ1 and Cγ2 grouped together, but their relation to the rodent Cγ genes was not resolved. The Cγ genes of the artiodactyls grouped together, while rabbit Cγ1 was not grouped with any other taxon. The phylogenetic relationships agree with previous hypotheses based on analysis of cDNA and protein similarity (Vernooij, 1993; Hein, 1994). Cγ4 appeared to have diverged early from the Cγ1 genes while Cγ2 is the product of a much more recent

duplication and inversion of Cγ1 (Vernooij, *et al.*, 1993). This hypothesis is also supported by the genomic sequence reported here.

It is clear that the divergence of Cγ1 and Cγ4 was ancient, but it was not possible to reach a conclusion as to whether the Cγ1 and Cγ4 genes were duplicated before or after the divergence of the primate and rodent lineages. Mice have four Cγ genes, whereas humans possess only two. Three of the mouse Cγ genes (Cγ1, Cγ2 and Cγ3) are 96% similar, while the Cγ4 gene shares just 65% homology with the others (Garman *et al.*, 1986). The human Cγ1 and Cγ2 are over 90% similar to each other (Lefranc *et al.*, 1986). The high homology of the mouse Cγ1, Cγ2 and Cγ3 genes suggest that they represent recent duplication products. The same is true for the human Cγ1 and Cγ2 genes. The mouse Cγ4 gene, however, is distinct and must be the product of an older duplication. Two hypotheses have been proposed for the divergence of the mouse Cγ4 gene. One possibility is that Cγ4 duplicated from Cγ1 before the divergence of the primate and rodent lineages and was subsequently deleted in the primate lineage. The other hypothesis is that the duplication occurred in the rodent lineage after the divergence of rodents and primates.

The lack of homology between the non-coding regions of Cγ2 and Cγ4 suggest that the two C gene have diverged from a very old duplication event. In comparison, the TCR Cα/Cδ region of human and mouse, separated for approximately 80 MYA, showed 71% of conservation in the intervening sequences between species (Koop and Hood, 1994). This suggests that Cγ4 diverged from Cγ1 and Cγ2 before the divergence of humans and mice. This conflicts with a previous phylogenetic analysis that placed the divergence of Cγ1 and Cγ4 at ~12 MYA previous estimate using the "Stationary Markov Clock" method on amino acid sequence alignments (Ciccarese *et al.*, 1997). Although there is variation in the rate of change in non-coding DNA at different genetic loci (Koop and Hood, 1994) and the Cα/Cδ region may be unique in its high level of conservation of non-coding DNA, the complete divergence of the introns of the Cγ2 and Cγ4 genes suggests that a divergence time much

greater than 12 MYA is likely. Comparison of the human and mouse TCR γ loci will provide a more complete picture of the evolutionary history of this locus.

The genomic sequence of the two V genes had been previously reported (Hayday *et al.*, 1985), as had the sequence of the genomic DNA lying between the two V genes (Kuziel *et al.*, 1984). As proposed by Vernooij *et al.* (1993), it seems likely that the Vγ1.1 gene segment was duplicated and inverted to give rise to the Vγ1.2 gene segment. The subsequent duplication of the γ2 cluster to give rise to the γ3 cluster would then account for the high homology of the Vγ1.1, 1.2 and 1.3 genes (Vernooij *et al.*, 1993). Vernooij *et al.* (1993) then suggested that the insertion of retroviral DNA between Vγ1.2 and Jγ2 after the formation of cluster γ3 led to the larger distance between Vγ1.2 and Jγ2. Having obtained the complete nucleotide sequence for this region, this hypothesis can be modified to implicate the insertion of two large LINEs as being responsible for increased V-J distance in the γ2 cluster.

Additional sequence data (Zhou *et al.*, unpublished data) from the 5' end of the mouse TCR γ locus also appears to support the theory proposed by Vernooij and co-workers (1993). The J gene segments and C genes of the γ1, γ2 and γ3 clusters all show strong conservation of both the coding and non-coding DNA, suggesting that these clusters are relatively recent duplication products. The region is also interspersed with a large number of full length LINEs and LINE fragments. The number of recently duplicated genes, LINEs and duplications within the LINEs themselves indicate that the TCR γ locus is an unstable and rapidly changing region of the mouse genome.

Gene duplication is thought to be a major driving force in evolution at the molecular level (Ohno, 1970). The TCRγ locus has undergone recent duplications yet retains a primitive organization of multiple clusters of gene segments with low combinatorial diversity. The γ2 cluster combines a V gene segment duplicated from the γ4 cluster with a J gene segment, C gene and enhancer duplicated from the γ1 cluster. The functional γ2

cluster was created from two separate duplication events, either one of which by itself would have created a pseudogene. Thus, one of the duplication products appears to have recruited the other to form a functional TCRγ cluster. The combination of gene segments duplicated in this manner may produce new clusters encoding receptors of novel specificity.

Whereas the sequence of the mouse TCRγ locus contained evidence of instability and recent change, the region has retained what appears to be a primitive organization with lower potential for combinatorial diversity than other TCR and Ig loci. When the combinatorial diversity of the TCRα subunit is compared with that of the TCRγ subunit, the difference is striking. From the genomic sequence data, it now appears that the number of Vγ gene segments contributing to functional TCRγ receptors is just six. There are three functional Jγ gene segments, and three functional Cγ genes, but each is contained in a separate cluster and there appears to be no sharing of gene segments between clusters (Garman, 1989). The combinatorial diversity of the mouse TCRγ locus is therefore limited to six. In contrast, the estimated 70 V gene segments and 61 J gene segments of the TCRα locus yield a combinatorial diversity of 4270. Thus, the potential combinatorial diversity of the mouse TCRγ locus is ~700x lower than that of the TCRα locus. If the TCR loci have evolved by duplicating an increasing number of V and J gene segments while reducing C gene repertoire to a single isotype (Matsunaga and Dahl, 1989; Janeway, 1992; Flajnik, 1994; Hein, 1994), the mouse TCRγ locus would appear to be an evolutionary relic. Hein (1994) proposed that the γδ T cells have a more primitive origin than the αβ T cells and Matsunaga and Dahl (1989) suggested that not only do the γδ T cells predate the αβ T cells in origin, but they also predate the B cells. It is therefore possible that the primitive organization of the mouse TCRγ locus is due to the maintenance of a primordial immune function, or functions, that continue to offer a selective advantage in the mouse. The low diversity clusters of gene segments suggests that a limited range of ligands of limited

diversity may be recognized by the γδ T cells. At the same time, however, the instability and evidence for recent duplications in the locus suggest that there may be some selection pressure to produce new receptor subunits.

The γδ TCR may recognize key ligands common to several important pathogens. These ligands, critical to the survival of the pathogens, would be under strong selection pressure both to maintain their function and to evade the host immune system. The combination of selection pressures might lead these ligands to achieve the maximum diversity possible within strict structural parameters required to maintain their critical function. Thus, the requirement to bind such ligands might give rise to an antigen receptor locus like that encoding the TCRγ subunit. This locus would need to encode receptors capable of recognizing a limited range of ligands, thus requiring only limited combinatorial diversity. The need to match the adaptations of pathogens might then account for the instability and gene duplications in the TCRγ locus, but because functional restrictions would limit the diversity of these ligands, the TCRγ locus would not need to adapt as quickly as other antigen receptor subunits.

4.5 Summary

The complete genomic sequence for 62 kb at the 3' end of the mouse TCR γ locus was obtained. Difficulty in obtaining sequence 5' of the γ2 cluster resulted from the fact that cosmid clones were found to be unstable and prone to rearrangement during growth. While frustrating, the instability of cosmids growing in *E. coli* may reflect the unstable nature of the corresponding region of the TCR γ locus. Sequencing this region using a more stable, low copy number vector may therefore provide interesting insight into an evolutionarily dynamic region of the mouse genome.

The genomic sequence data obtained in this study represents the first contiguous genomic sequence to cover two mouse TCR γ gene clusters. This made possible the discovery of a putative γ4 enhancer sequence, a novel repeat element, and a map of the

genome-wide repeat elements located within the locus. Analysis of internal homology within the genomic sequence was used to determine the location and extent of duplicated regions within the locus and comparison of the two TCR gene clusters within the sequence indicated that the two C genes were the products of an ancient duplication event, whereas the V genes had been duplicated much more recently.

Based on the primitive organization and low combinatorial diversity of this locus, it was suggested that the mouse γδ T cells may occupy a functional niche that has been maintained from the primordial vertebrate immune system. Further studies of the mouse γδ T cells may therefore help to understand the origins of the vertebrate immune system and aid in understanding the current role of the γδ T cell in humans.

Combining this data with that of other researchers will eventually provide a complete sequence for the entire TCR γ locus. From this complete sequence it should be possible to study in detail the evolutionary events recorded in the intervening DNA through comparisons of this sequence with both itself and the TCR loci of other organisms.

REFERENCES

Allison, J.P., and Havran, W.L. 1991. The immunobiology of T cells with invariant γδ antigen receptors. Annual Review of Immunology 9:679-705.

Angiolillo, A.L., Lamoyi, E., Bernstein, K.E. and Mage, R.G. 1985. Identification of genes for the constant region of rabbit T-cell receptor β chains. Proceedings of the National Academy of Sciences of the United States of America 82: 4498-4502.

Asarnow, D.M., Kuziel, W.A., Bonyhadi, M., Tigelaar, R.E., Tucker, P.W. and Allison, J.P. 1988. Limited diversity of γδ antigen receptor genes of Thy-1+ dendritic epidermal cells. Cell 55: 837-847.

Austyn, J.M. And Wood, K.J. 1993. *Principles of Cellular and Molecular Immunology.* Oxford Univesity Press, Oxford.

Bandeira, A., Itohara, S., Bonneville, M., Burlen-Defranoux, O., Mota-Santos, T., Coutinho, A. and Tonegawa, S. 1991. Extrathymic origin of intestinal intraepithelial lymphocytes bearing T-cell antigen receptor γδ. Proceedings of the National Academy of Sciences of the United States of America 88: 43-47.

Beckman, E.M., Porcelli, S.A., Morita, C.T., Behar, S.M., Furlong, S.T. and Brenner, M.B. 1994. Recognition of a lipid antigen by CD1-restricted αβ+ T cells. Nature 372: 691-694.

Beckman, EM. and Brenner M.B. 1995. MHC class I-like, class II-like and CD1 molecules: distinct roles in immunity. Immunology Today 16: 349-352.

Behr, C. and Dubois, P. 1992. Preferential expansion of Vγ9 Vδ2 T cells following stimulation of peripheral blood lymphocytes with extracts of *Plasmodium falciparum.* International Immunology 4: 361-366.

Bodenteich, S., Chissoe, S., Wang, Y.-F. and Roe, B.A. 1994. Shotgun cloning as the strategy of choice to generate templates for high-throughput dideoxynucleotide sequencing. *In* Adams, M.D., Fields, C., and Venter, J.C. (eds.) *Automated DNA Sequencing and Analysis.* Academic Press, London. pp. 42-50.

Born, W., Hall, L., Dallas, A., Boymel, J., Shinnick, T., Young, D., Brennan, P., and O'Brien, R. 1990. Recognition of a peptide antigen by heat shock--reactive γδ T lymphocytes. Science 249: 67-69.

Bluestone, J.A., Cron, R.Q., Barrett, T.A., Houlden, B., Sperling, A.I., Dent, A., Hedrick, S., Rellahan, B. and Matis, L.A. 1991. Repertoire development and ligand specificity of murine TCR γδ cells. Immunological Reviews 120: 5-33.

Bonneville, M., Janeway, C.A., Ito, K., Haser, W., Ishida, I., Nakanishi, N. and Tonegawa, S. 1988. Intestinal intraepithelial lymphocytes are a distinct set of γδ T cells. Nature 336: 479-481.

Bougueleret, L. and Claverie, J.M. 1987. Variability analysis of the human and mouse T-cell receptor β chains. Immunogenetics 26: 304-308.

Bürk M.R., Mori, L. and De Libero, G. 1995. Human Vγ9-Vδ2 cells are stimulated in a cross-reactive fashion by a variety of phosphorylated metabolites. European Journal of Immunology 25: 2052-2058.

Carding, S.R., Kyes, S., Jenkinson, E.J., Kingston, R., Bottomly, K., Owen, J.J., and Hayday, A.C. 1990. Developmentally regulated fetal thymic and extrathymic T-cell receptor γδ gene expression. Genes and Development 4: 1304-1315.

Chen, Z., Font, M.P., Loiseau, P., Bories, J.C., Degos, L., Lefranc, M.P., and Sigaux, F. 1988. The human T-cell Vγ gene locus: cloning of new segments and study of Vγ rearrangements in neoplastic T and B cells. Blood 72: 776-783.

Cheng, S., Fockler, C., Barnes, W., and Higuchi, R. 1994. Effective amplification of long targets from cloned inserts and human genomic DNA. Proceedings of the National Academy of Sciences of the United States of America 91: 5695-5699.

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

Ciccarese, S., Lanave, C., and Saccone, C. 1997. Evolution of T-cell receptor γ and δ constant region and other T-cell-related proteins in the human-rodent-artiodactyl triplet. Genetics 145: 409-419.

Constant, P., Davodeau, F., Peyrat, M.A., Poquet, Y., Puzo, G., Bonneville, M. and Fournie, J.J. 1994. Stimulation of human γδ T cells by nonpeptidic mycobacterial ligands. Science 264: 267-270.

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

De Libero, G., Casorati, G., Giachino, C., Carbonara, C., Migone, N., Matzinger, P., and Lanzavecchia, A. 1991. Selection by two powerful antigens may account for the presence of the major population of human peripheral γ/δ T cells. Journal of Experimental Medicine 173: 1311-1322.

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

Ezquerra, A., Cron, R.Q., McConnell, T.J. Valas, R.B., Bluestone, J.A. Coligan, J.E. 1990. TCR δ gene expression and diversity in the mouse spleen. Journal of Immunology 145: 1311-1317.

Fitch, D.H., Bailey, W.J., Tagle, D.A., Goodman, M., Sieu, L., and Slightom, J.L. 1991. Duplication of the γ-globin gene mediated by L1 long interspersed repetitive elements in an early ancestor of simian primates. Proceedings of the National Academy of Sciences U.S.A. 88: 7396-7400.

Flajnik, M.F., Canel, C., Kramer, J., and Kasahara, M. 1991a. Evolution of the major histocompatibility complex: molecular cloning of major histocompatibility complex class I from the amphibian *Xenopus*. Proceedings of the National Academy of Sciences of the United States of America 88: 537-541.

Flajnik, M.F., Canel, C., Kramer, J. and Kasahara, M. 1991b. Which came first, MHC class I or class II? Immunogenetics 33: 295-300.

Flajnik, M.F. 1994. Primitive vertebrate immunity: what is the evolutionary derivative of molecules that define the adaptive immune system? Ciba Foundation Symposiums 186: 224-232.

Flynn, G.C., Chappell, T.G. and Rothman, J.E. 1989. Peptide binding and release by proteins implicated as catalysts of protein assembly. Science 245: 385-390.

Forster, A., Huck, S., Ghanem, N., and Lefranc, M.P., and Rabbitts, T.H. 1987. New subgroups in the human T cell rearranging V γ gene locus. EMBO Journal 6: 1945-1950.

Garman, R.D., Doherty P.J. and Raulet D.H. 1986. Diversity, rearrangement, and expression of murine T cell γ genes. Cell 45: 733-742.

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

Grossberger, D., Marcuz, A., Fichtel, A., Dudler, L. and Hein, W.R. 1993. Sequence analysis of sheep T-cell receptor β chains. Immunogenetics 37: 222-226.

Groettrup, M., Ungewiss, K., Azogui, O., Palacios, R., Owen, M.J., Hayday, A.C. and von Boehmer, H. 1993. A novel disulfide-linked heterodimer on pre-T cells consists of the T cell receptor β chain and a 33 kd glycoprotein. Cell 75: 283-294.

Guo, Y., Ziegler, H.K., Safley, S.A., Niesel, D.W., Vaidya, S. and Klimpel, G.R. 1995. Human T-cell recognition of Listeria monocytogenes: recognition of listeriolysin O by TcR $\alpha\beta$+ and TcR $\gamma\delta$+ T cells. Infection and Immunity 63: 2288-2294.

Guy-Grand, D., Cerf-Bensussan, N., Malissen, B., Malassis-Seris,M., Briottet, C. and Vassalli, P. 1991. Two gut intraepithelial CD8+ lymphocyte populations with different T cell receptors: a role for the gut epithelium in T cell differentiation. Journal of Experimental Medicine 173: 471-481.

Haars, R., Kronenberg, M., Gallatin, W.M., Weissman, I.L., Owen, F.L. and Hood, L. 1986. Rearrangement and expression of T cell antigen receptor and γ genes during thymic development. Journal of Experimental Medicine 164: 1-24.

Haas, W., Pereira, P., and Tonegawa, S. 1993. γ/δ cells. Annual Review of Immunology 11: 637-685.

Havran, W.L. and Allison, J.P. 1988. Developmentally ordered appearance of thymocytes expressing different T-cell antigen receptors. Nature 335: 443-445.

Hayday, A.C., Saito, H., Gillies, S.D., Kranz, D.M., Tanigawa, G., Eisen H.N. and Tonegawa, S. 1985. Structure, organization, and somatic rearrangement of T cell γ genes. Cell 40: 259-269.

Hedrick, S.M., Nielsen, E.A., Kavaler, J., Cohen, D.I., and Davis, M.M. 1984. Sequence relationships between putative T-cell receptor polypeptides and immunoglobulins. Nature 308: 153-158.

Heilig J.S. and Tonegawa, S. 1986. Diversity of murine γ genes and expression in fetal and adult T lymphocytes. Nature 322: 836-840.

Hein, W.R., Dudler, L., Marcuz, A. and Grossberger, D. 1990. Molecular cloning of sheep T cell receptor γ and δ chain constant regions: unusual primary structure of γ chain hinge segments. European Journal of Immunology 20: 1795-1804.

Hein, W.R. and Dudler, L. 1993. Divergent evolution of T cell repertoires: extensive diversity and developmentally regulated expression of the sheep γδ T cell receptor. EMBO Journal 12: 715-724.

Hein, W.R. 1994. Structural and functional evolution of the extracellular regions of T cell receptors. Seminars in Immunology 6: 361-372.

Holoshitz, J., Koning, F., Coligan, J.E., De Bruyn, J. and Strober, S. 1989. Isolation of CD4- CD8- mycobacteria-reactive T lymphocyte clones from rheumatoid arthritis synovial fluid. Nature 339: 226-229.

Huang, X. and Miller, W. 1991. A time-efficient, linear-space local similarity algorithm. Advances in Applied Mathematics 12: 337-358.

Huck, S., and Lefranc, M.P. 1987. Rearrangements to the JP1, JP and JP2 segments in the human T-cell rearranging γ gene (TRG γ) locus. FEBS Letters 224: 291-296.

Imani, F. and Soloski, M.J. 1991. Heat shock proteins can regulate expression of the Tla region-encoded class Ib molecule Qa-1. Proceedings of the National Academy of Sciences of the United States of America 88: 10475-10479.

Ishiguro, N., Aida, Y., Shinagawa, T. and Shinagawa, M. 1993. Molecular structures of cattle T cell receptor γ and δ chains predominantly expressed on peripheral blood lymphocytes. Immunogenetics 38: 437-443.

Isono, T., Kim, C.J. and Seto, A. 1995. Sequence and diversity of rabbit T-cell receptor γ chain genes. Immunogenetics 41: 295-300.

Ito, K., Bonneville, M., Takagaki, Y., Nakanishi, N., Kanagawa, O., Krecko, E.G. and Tonegawa, S. 1989. Different γδ T-cell receptors are expressed on thymocytes at different stages of development. Proceedings of the National Academy of Sciences of the United States of America 86: 631-635.

Ito, K., Tsunoda, M., Watanabe, K., Ito, K., Kashiwagi, N. and Obata, F. 1993. Isolation and sequence analysis of cDNA for the dog T-cell receptor Tcr α and Tcr β chains. Immunogenetics 38: 60-63.

Itohara, S., Nakanishi, N., Kanagawa, O., Kubo, R. and Tonegawa, S. 1989. Monoclonal antibodies specific to native murine T-cell receptor $\gamma\delta$: analysis of $\gamma\delta$ T cells during thymic ontogeny and in peripheral lymphoid organs. Proceedings of the National Academy of Sciences of the United States of America 86: 5094-5098.

Itohara, S., Farr, A.G., Lafaille, J.J., Bonneville, M., Takagaki, Y., Haas, W. & Tonegawa, S. 1990. Homing of a $\gamma\delta$ thymocyte subset with homogeneous T-cell receptors to mucosal epithelia. Nature 343: 754-757.

Iwamoto, A., Rupp, F., Ohashi, P.S., Walker, C.L., Pircher, H., Joho, R., Hengartner, H. and Mak, T.W. 1986. T cell-specific γ genes in C57BL/10 mice. Sequence and expression of new constant and variable region genes. Journal of Experimental Medicine 163: 1203-1212.

Janeway, C.A. Jr. 1992. The immune system evolved to discriminate infectious nonself from noninfectious self. Immunology Today 13: 11-16.

Jones, N., Leiden, J., Dialynas, D., Fraser, J., Clabby, M., Kishimoto, T., Strominger, J.L., Andrews, D., Lane, W., and Woody, J. 1985. Partial primary structure of the alpha and beta chains of human tumor T-cell receptors. Science 227: 311-314.

Jouvin-Marche, E., Heller, M. and Rudikoff, S. 1986. Gene correction in the evolution of the T cell receptor β chain. Journal of Experimental Medicine 164: 2083-2088.

Kabelitz, D., Bender, A., Schondelmaier, S., Schoel, B., and Kaufmann, S.H. 1990. A large fraction of human peripheral blood γ/δ + T cells is activated by *Mycobacterium tuberculosis* but not by its 65-kD heat shock protein. Journal of Experimental Medicine 171: 667-679.

Kaliyaperumal, A., Falchetto, R., Cox, A., Dick, R., Shabanowitz, J., Chien, Y.H.,Matis, L., Hunt, D.F. and Bluestone, J.A. 1995. Functional expression and recognition of nonclassical MHC class I T10b is not peptide-dependent. Journal of Immunology 155: 2379-2386.

Kasahara, M., Vazquez, M., Sato, K., McKinney, E.C. and Flajnik, M.F. 1992. Evolution of the major histocompatibility complex: isolation of class II A cDNA clones from the cartilaginous fish. Proceedings of the National Academy of Sciences of the United States of America 89: 6688-6692.

Kaufmann, S.H. 1990. Heat shock proteins and the immune response. Immunology Today 11: 129-136.

Kaufmann, S.H. 1996. γ/δ and other unconventional T lymphocytes: what do they see and what do they do? Proceedings of the National Academy of Sciences of the United States of America 93: 2272-2279.

Kim, C.J., Isono, T., Tomoyoshi, T. and Seto, A. 1994. Expression of T-cell receptor γ/δ chain genes in a rabbit killer T-cell line. Immunogenetics 39: 418-422.

Kinebuchi, M., Kikuchi, K., and Matsuura, A. 1994. Rat T-cell receptor γ chain sequences: identification of conserved gene segments and a unique chimeric constant region gene, C4L. Immunogenetics 40: 449-455.

Kishihara, K., Yoshikai, Y., Matsuzaki, G., Tomooka, S., and Nomoto, K. 1988. "Radioresistant" intrathymic T cell precursors express T cell receptor Cγ4- and Cδ-specific gene messages. European Journal of Immunology 18: 841-847.

Klotz, J.L., Barth, R.K., Kiser, G.L., Hood, L.E., and Kronenberg, M. 1989. Restriction fragment length polymorphisms of the mouse T-cell receptor gene families. Immunogenetics 29: 191-201.

Kokubu, F., Litman, R., Shamblott, M.J., Hinds, K. and Litman, G.W. 1988. Diverse organization of immunoglobulin VH gene loci in a primitive vertebrate. EMBO Journal 7: 3413-3422.

Koop, B.F. and Hood, L. 1994. Striking sequence similarity over almost 100 kilobases of human and mouse T-cell receptor DNA. Nature Genetics 7: 48-53.

Kozbor, D., Trinchieri, G., Monos, D.S., Isobe, M., Russo, G., Haney, J.A., Zmijewski, C. and Croce, C.M. 1989. Human TCR-γ+/δ+, CD8+ T lymphocytes recognize tetanus toxoid in an MHC-restricted fashion. Journal of Experimental Medicine 169: 1847-1851.

Krangel, M.S., Band, H., Hata, S., McLean, J. and Brenner, M.B. 1987. Structurally divergent human T cell receptor γ proteins encoded by distinct Cγ genes. Science 237: 64-67.

Kuziel, W.A., Kienker, L.J., and Tucker, P.W. 1994. Physical linkage of mouse Tcrg-V genes. I. Evolutionary and regulatory features of the Vg1.1-Vg1.2 intergenic region. Immunogenetics 39: 296-297.

Kyes, S., Carew, E., Carding, S.R., Janeway, C.A. and Hayday, A. 1989. Diversity in T-cell receptor γ gene usage in intestinal epithelium. Proceedings of the National Academy of Sciences of the United States of America 86: 5527-5531.

Lai, E., Wilson, R.K. and Hood, L.E. 1989. Physical maps of the mouse and human immunoglobulin-like loci. Advances in Immunology 46: 1-59.

Lafaille, J.J., DeCloux, A., Bonneville, M., Takagaki, Y. and Tonegawa, S. 1989. Junctional sequences of T cell receptor γδ genes: implications for γδ T cell lineages and for a novel intermediate of V-(D)-J joining. Cell 59: 859-870.

Lefranc, M.P., Forster, A. and Rabbitts, T.H. 1986. Genetic polymorphism and exon changes of the constant regions of the human T-cell rearranging gene γ. Proceedings of the National Academy of Sciences of the United States of America 83: 9596-9600.

Lefranc, M.P. and Rabbitts, T.H. 1990. Genetic organization of the human T-cell receptor γ and δ loci. Research in Immunology 141: 565-577.

Lefrancois, L., LeCorre, R., Mayo, J., Bluestone, J.A. and Goodman, T. 1990. Extrathymic selection of TCR γδ+ T cells by class II major histocompatibility complex molecules. Cell 63: 333-340.

Lieber, M.R. 1991. Site-specific recombination in the immune system. FASEB Journal 5: 2934-2944.

Litman, G.W., Wang, A.C., Fudenberg, H.H., and Good, R.A. 1971. N-terminal amino-acid sequence of African lungfish immunoglobulin light chains. Proceedings of the National Academy of Sciences of the United States of America 68:2321-2324.

Loh, E.Y., Lanier, L.L., Turck, C.W., Littman, D.R., Davis, M.M., Chien, Y.H. and Weiss, A. 1987. Identification and sequence of a fourth human T cell antigen receptor chain. Nature 330: 569-572.

Malissen, M., Minard, K., Mjolsness, S., Kronenberg, M., Goverman, J., Hunkapiller, T., Prystowsky, M.B., Yoshikai, Y., Fitch, F., Mak, T.W. and Hood, L. 1984. Mouse T cell antigen receptor: structure and organization of constant and joining gene segments encoding the beta polypeptide. Cell 37: 1101-1110.

Marche, P.N. and Kindt, T.J. 1986. A variable region gene subfamily encoding T cell receptor β-chains is selectively conserved among mammals. Journal of Immunology 137: 1729-1734.

Marchalonis, J.J., Decker, J.M., DeLuca, D., Moseley, J.M., Smith, P. and Warr, G.W. 1977. Lymphocyte surface immunoglobulins: evolutionary origins and involvement in activation. Cold Spring Harbor Symposia on Quantitative Biology 41: 261-273.

Matis, L.A., Fry, A.M., Cron, R.Q., Cotterman, M.M., Dick, R.F. and Bluestone, J.A. 1989. Structure and specificity of a class II MHC alloreactive γδ a T cell receptor heterodimer. Science 245: 746-749.

Matsunaga, T. and Dahl, U. 1989. What was wrong with the T-cell receptor γ/δ heterodimer? Divergence of the T-cell receptor α/δ and γ/δ heterodimers. Scandinavian Journal of Immunology 30: 511-517.

McCormack, W.T., Tjoelker, L.W., Stella, G., Postema, C.E. and Thompson, C.B. 1991. Chicken T-cell receptor β-chain diversity: an evolutionarily conserved D β-encoded glycine turn within the hypervariable CDR3 domain. Proceedings of the National Academy of Sciences of the United States of America 88: 7699-7703.

Melchers, F., Karasuyama, H., Haasner, D., Bauer, S., Kudo, A., Sakaguchi, N., Jameson, B. and Rolink, A. 1993. The surrogate light chain in B-cell development. Immunology Today 14: 60-68.

Modlin, R.L., Pirmez, C., Hofman, F.M., Torigian, V., Uyemura, K., Rea, T.H., Bloom, B.R., and Brenner, M.B. 1989. Lymphocytes bearing antigen-specific γδ T-cell receptors accumulate in human infectious disease lesions. Nature 339: 544-548.

Morita, C.T., Beckman, E.M., Bukowski, J.F., Tanaka, Y., Band, H., Bloom, B.R., Golan, D.E. and Brenner, M.B. 1995. Direct presentation of nonpeptide prenyl pyrophosphate antigens to human γδ T cells. Immunity 3: 495-507.

Mosley, R.L., Styre, D., and Klein, J.R. 1990. Differentiation and functional maturation of bone marrow-derived intestinal epithelial T cells expressing membrane T cell receptor in athymic radiation chimeras. Journal of Immunology 145:1369-1375.

Munk, M.E., Gatrill, A.J., and Kaufmann, S.H. 1990. Target cell lysis and IL-2 secretion by γ/δ T lymphocytes after activation with bacteria. Journal of Immunology 145: 2434-2439.

O'Brien, R.L., and Born, W. 1991. Heat shock proteins as antigens for γ δ T cells. Seminars in Immunology 3:81-87.

O'Brien, R.L., Happ, M.P., Dallas, A., Cranfill, R., Hall, L., Lang, J., Fu, Y.X., Kubo, R., and Born, W. 1991. Recognition of a single hsp-60 epitope by an entire subset of γ δ T lymphocytes. Immunological Reviews 121:155-170.

O'Brien, R.L., Fu, Y.X., Cranfill, R., Dallas, A., Ellis, C., Reardon, C., Lang, J., Carding, S.R., Kubo, R., Born, W. 1992. Heat shock protein Hsp60-reactive γδ cells: a large, diversified T-lymphocyte subset with highly focused specificity. Proceedings of the National Academy of Sciences of the United States of America 89: 4348-4352.

Ohno, S. 1970. *Evolution by Gene Duplication*. Springer-Verlag, Berlin.

Pelham, H. 1988. Heat-shock proteins. Coming in from the cold. Nature 332: 776-777.

Pfeffer, K., Schoel, B., Gulle, H., Kaufmann, S.H. and Wagner, H. 1990. Primary responses of human T cells to mycobacteria: a frequent set of γ/δ T cells are stimulated by protease-resistant ligands. European Journal of Immunology 20: 1175-1179.

Porcelli, S., Morita, C.T., and Brenner M.B. 1992. CD1b restricts the response of human CD4-8- T lymphocytes to a microbial antigen. Nature 360: 593-597.

Porcelli, S.A. 1995. The CD1 family: a third lineage of antigen-presenting molecules. Advances in Immunology 59: 1-98.

Povenelli, C.M., and Gibbs, R.A. 1993. Large-scale sequencing library production: an adaptor-based strategy. Analytical Biochemistry 210: 16-26.

Raulet, D.H. 1989. The structure, function, and molecular genetics of the γ/δ T cell receptor. Annual Reviews in Immunology 7: 175-207.

Raulet, D.H. 1994. MHC class I-deficient mice. Advances in Immunology 55: 381-421.

Rock, E.P., Sibbald, P.R., Davis, M.M. and Chien, Y.H. 1994. CDR3 length in antigen-specific immune receptors. Journal of Experimental Medicine 179: 323-328.

Roitt, I. 1991. *Essential Immunology*. 7th ed. Blackwell Scientific Publications, London.

Rowen, L., Koop, B.F., and Hood, L. 1996. The complete 685-kilobase DNA sequence of the human β T cell receptor locus. Science 272: 1755-1762.

Sambrook, J., Fritsch, E.F. and Maniatis, T. 1989. *Molecular Cloning: a Laboratory Manual*. 2nd Ed. Cold Spring Harbor Laboratory Press.

Schiffer, M., Wu, T.T. and Kabat, E.A. 1986. Subgroups of variable region genes of β chains of T-cell receptors for antigen. Proceedings of the National Academy of Sciences of the United States of America 83: 4461-4463.

Schiffer, M., Kabat, E.A. and Wu, T.T. 1992. Subgroups of Tcr α chains and correlation with T-cell function. Immunogenetics 35: 224-234.

Schild, H., Mavaddat, N., Litzenberger, C., Ehrich, E.W., Davis, M.M., Bluestone, J.A., Matis, L., Draper, R.K., and Chien, Y.H. 1994. The nature of major histocompatibility complex recognition by γδ T cells. Cell 76: 29-37.

Schoel, B., Sprenger, S. and Kaufmann, S.H. 1994. Phosphate is essential for stimulation of Vγ9 Vδ2 T lymphocytes by mycobacterial low molecular weight ligand. European Journal of Immunology 24: 1886-1892.

Sieling, P.A., Chatterjee, D., Porcelli, S.A., Prigozy, T.I., Mazzaccaro, R.J., Soriano, T., Bloom, B.R., Brenner, M.B., Kronenberg, M., Brennan, P.J., and Modlin, R.L. 1995. CD1-restricted T cell recognition of microbial lipoglycan antigens. Science 269: 227-230.

Soloski, M.J., DeCloux, A., Aldrich, C.J., and Forman, J. 1995. Structural and functional characteristics of the class IB molecule, Qa-1. Immunological Reviews 147: 67-89.

Sonnhammer, E.L.L. and Durbin, R. 1996. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene 167: GC1-10.

Spencer, D.M., Hsiang, Y.H., Goldman, J.P. and Raulet, D.H. 1991. Identification of a T-cell-specific transcriptional enhancer located 3' of Cγ1 in the murine T-cell receptor γ locus. Proceedings of the National Academy of Sciences of the United States of America 88: 800-804.

Takagaki, Y., Nakanishi, N.,Ishida, I., Kanagawa, O. and Tonegawa, S. 1989a. T cell receptor-γ and -δ genes preferentially utilized by adult thymocytes for the surface expression. Journal of Immunology 142: 2112-2121.

Takagaki, Y., DeCloux, A., Bonneville, M. and Tonegawa, S. 1989b. Diversity of γδ T-cell receptors on murine intestinal intra-epithelial lymphocytes. Nature 339: 712-714.

Takeuchi, N., Ishiguro, N. and Shinagawa, M. 1992. Molecular cloning and sequence analysis of bovine T-cell receptor γ and δ chain genes. Immunogenetics 35: 89-96.

Tanaka, A., Ishiguro, N. and Shinagawa, M. 1990. Sequence and diversity of bovine T-cell receptor β-chain genes. Immunogenetics 32: 263-271.

Tanaka, Y., Sano, S., Nieves, E., De Libero, G., Rosa, D., Modlin, R.L., Brenner, M.B., Bloom, B.R. and Morita, C.T. 1994. Nonpeptide ligands for human γδ T cells. Proceedings of the National Academy of Sciences of the United States of America 91: 8175-8179.

Tanaka, Y., Morita, C.T., Tanaka, Y., Nieves, E., Brenner, M.B. and Bloom, B.R. 1995. Natural and synthetic non-peptide antigens recognized by human γδ T cells. Nature 375: 155-158.

Thome, A., Saalmuller, A. and Pfaff, E. 1993. Molecular cloning of porcine T cell receptor α, β, γ and δ chains using polymerase chain reaction fragments of the constant regions. European Journal of Immunology 23: 1005-1010.

Tonegawa, S. 1983. Somatic generation of antibody diversity. Nature 302: 575-581.

Traunecker, A., Oliveri, F., Allen, N. and Karjalainen, K. 1986. Normal T cell development is possible without 'functional' γ chain genes. EMBO Journal 5: 1589-1593.

Uberbacher, E.C. and Mural, R.J. 1991. Locating protein coding regions in human dna sequences using a multiple sensor-neural network approach. Proceedings of the National Academy of Sciences of the United States of America 88: 11261-11265.

Vernooij, B.T. 1992. *The Mouse T Cell Receptor γ Genes*. Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy. California Institute of Technology, Pasadena.

Vernooij, B.T., Lenstra, J.A., Wang, K. and Hood, L. 1993. Organization of the murine T-cell receptor γ locus. Genomics 17: 566-574.

Vidovic, D., Roglic, M., McKune, K., Guerder, S., MacKay, C. & Dembic, Z. 1989. Qa-1 restricted recognition of foreign antigen by a γδ T-cell hybridoma. Nature 340: 646-650.

Williams, A.F. 1987. A year in the life of the immunoglobulin superfamily. Immunology Today 8: 293-304.

Williams, C.B., Blankenhorn, E.P., Byrd, K.E., Levinson, G. and Gutman, G.A. 1991. Organization and nucleotide sequence of the rat T cell receptor β-chain complex. Journal of Immunology 146: 4406-4413.

Weintraub, B.C., Jackson, M.R. and Hedrick, S.M. 1994. γδ T cells can recognize nonclassical MHC in the absence of conventional antigenic peptides. Journal of Immunology 153: 3051-3058.

Yanagi, Y., Yoshikai, Y., Leggett, K., Clark, S.P., Aleksander, I. and Mak, T.W. 1984. A human T cell-specific cDNA clone encodes a protein having extensive homology to immunoglobulin chains. Nature 308: 145-149.

Zhang, X.M., Huck, S., Dariavach, P., Spilliaert, R., Arnason, A., Soua, Z., Lefranc, G. and Lefranc, M.P. 1995. Isolation of T cell receptor γ/δ chain constant genes from minke whale. Experimental and Clinical Immunogenetics 12: 245-252.

Zhu, X., Zhao, X., Burkholder, W.F., Gragerov, A., Ogata, C.M., Gottesman, M.E., and Hendrickson, W.A. 1996. Structural analysis of substrate binding by the molecular chaperone DnaK. Science 272: 1606-1614.

VITA

Surname: Parlee Given Names: Michael John

Place of Birth: Sidney, British Columbia, Canada

Educational Institutions Attended:

University of Victoria	1988 to 1993

Degrees Awarded:

B.Sc.	University of Victoria	1993

Honours and Awards:

Dr. Julius F. Schleicher Graduate Student Scholarship	1995

PARTIAL COPYRIGHT LICENSE

I hereby grant the right to lend my thesis to users of the University of Victoria Library, and to make single copies only for such users or in response to a request from the library of any other university, or similar institution, on its behalf or for one of its users. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by me or a member of the University designated by me. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Title of Thesis:

Genomic Sequence and Evolutionary Analysis of the 3' End of the Murine T Cell Receptor γ Locus

Author [redacted]
Michael John Parlee
September 30, 1997