The Mouse T Cell Receptor Gamma Genes.

Thesis by

Bernardus Theodorus Maria Vernooij

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1992

(Defended May 18, 1992)

ABSTRACT.

Murine T cells express either of 2 antigen receptors on their surface: αß or γδ T cell receptors. The γδ T cell population contains subsets which show tissue specific localization, invariant T cell receptors and/or specificity for stress antigens. This makes these T cells unlike αß T cells.

This thesis describes the genomic organization of the entire mouse T cell receptor gamma locus. It contains 4 clusters of gene segments, each with a C, a J and 1 to 4 V gene segments. Compared to other T cell receptor and immunoglobulin loci, this is an unusual organization. The Cγ2 cluster is in an orientation that is opposite to that of all other clusters.

Two new γ enhancer-like elements were identified in the locus. Also shown is that the hinge region of Cγ4 is encoded by at least 2 exons. This is similar to the gene organization of the human Cγ2 gene segment, and different from the other mouse and human Cγ gene segments.

Sequence comparison of the T cell receptor γ gene segments of various mammals reveals structural conservation during evolution. The C region is most conserved, except in the hinge region. This subdomain is variable in length and in sequence. The extracellular domain is well conserved and contains amino acid residues which are also conserved in the other T cell receptor and immunoglobulin proteins.

The Vγ gene segments are less well conserved, but several amino acid residues are found which are (nearly) invariant. During evolution, the 2 studied mammals each appear to have lost certain V gene segments relative to a hypothetical ancestor.

Acknowledgements.

I most want to thank my wife, Nicky. You have always encouraged me and without you I would not be writing my thesis. Thank you from the bottom of my heart. I wish I could share the degree with you. You deserve it.

I also want to thank my parents for what they have done for me. They encouraged me throughout my life and gave me a set of values for life.

In the lab, I have had help from many. Ulf was and still is a good friend and was great to talk to, about anything and everything. Kai and Chia-Lam, it was nice working with you. A thank you goes to Kai for teaching me some of the finer point of mapping. In the final stages of this endeavour, the encouragement from and proofreading of my verse by Mike, Ben and Kai was helpful. The friendliness from the people in the lab was always encouraging. Thank you Ria, Bertha, Jessie and Karen, not only for keeping the show going, but also for being yourselves. A thank you is also due to all the less visible people who allowed us to concentrate on working, including Patty, Mary, Isabella, Bill, Sue and Karen.

Also, a special thank you goes to all members of my committee, Lee Hood, Elliot Meyerowitz, Ellen Rothenberg, Mel Simon, Pamela Bjorkman and Scott Emr early on. Having frequent committee meetings, especially in the end, has helped me keep on track. I strongly urge you to keep doing this for students

who appear to become part of the furniture. A special thank you goes to Lee for accepting me into his lab and for uniting me with Nicky.

Table of contents.

Chapter I.

INTRODUCTION.

The immune system.

The mammalian immune system protects the host from infections by neutralizing and eliminating pathogens. Several cell types contribute to the immune system, which can be divided into nonspecific and antigen (Ag) specific cells.

Macrophages and monocytes are nonspecific effector cells that phagocytose and neutralize foreign material.

Lymphocytes on the other hand are antigen specific effector cells. These include the B and T cells. B cells can secrete their antigen receptor, the immunoglobulin (Ig), into the extracellular fluid. There it binds specifically to the pathogenic substance (free antigen). Following this binding, the pathogen is neutralized by phagocytic cells, or by complement mediated cell lysis.

Unlike B cells, T cells can not secrete their receptor molecule. The antigen specific T cell receptor (TCR) remains anchored in the cell membrane. Like the antibodies (Abs), it is highly specific in its recognition. The TCR binds to antigen fragments present on the surface of other cells. Ag binding activates the T cells. Activated helper T cells (Th) respond to binding of antigen by producing growth and differentiation factors, which stimulate ('help') other cells of the immune system. Cytotoxic T cells (Tc) that recognize virus Ag on infected cells on the other hand kill

the infected cell.

There are 2 types of T cells, which can be distinguished by their TCR: αß T cells have a TCR composed of an α and a ß chain, whereas the γδ cells have a TCR made up of a γ and a δ chain.

The basic structure of the receptors (Ig, αß and γδ TCR) is very similar. They all consist of 2 chains, each with a constant and a variable part. The recognition specificity resides in the combined variable parts. Each cell makes chains with unique variable parts and thus, each cell produces a single type of receptor, which is highly specific for a particular antigen.

Diversity.

The immune system must be capable of recognizing every foreign antigen which may enter the body. Hence, the receptors on the B and T cells must be capable of recognizing an enormous diversity of antigens. The diversity in the receptors is generated by DNA rearrangements in individual T and B cells that juxtapose 2 or 3 gene segments from a large number of such elements, thereby creating a gene which encodes the variable domain of the receptor.

The Ig light (IgL) chain and the TCR α and γ loci contain multiple V and J gene segments, which are physically separated on the DNA. Figure 1 shows how one Vγ and one Jγ

gene segment are joined in a DNA rearrangement process which generates a Vγ gene. The rearrangement process is guided by DNA rearrangement signals 3' to the V and 5' to the J segments. These signals consist of conserved 7-mers and 9-mers and a nonconserved 12 or 23 basepair (bp) spacer. In the rearrangement process, these signals and the intervening DNA are excised.

The Ig heavy (IgH) chain and the TCR ß and δ loci have multiple V and J gene segments as well as multiple D elements, which are located between the V and J elements. In these loci, a J segment is joined to a D, which is then joined to a V, generating a V gene.

Additional diversity arises from imprecise joining of V, D and J DNA segments, due to removal of bases from the ends of the gene segments and/or insertion of random bases (N nucleotides) at the joints before ligation of the DNA ends. As a result, the junctions which result from a particular V to J (or V-D or D-J) rearrangement can be very heterogeneous. As the junctions encode part of the antigen binding site, this process creates useful additional receptor diversity.

If the resulting V gene is in the proper translational reading frame (i.e., functional), it will encode one chain of the receptor. Finally, each chain which is thus encoded can pair with another chain (α and ß, γ and δ, IgH and IgL)

also created by the mechanism outlined above (Tonegawa, 1983; Davis and Bjorkman, 1988; Lafaille *et al.*, 1989).

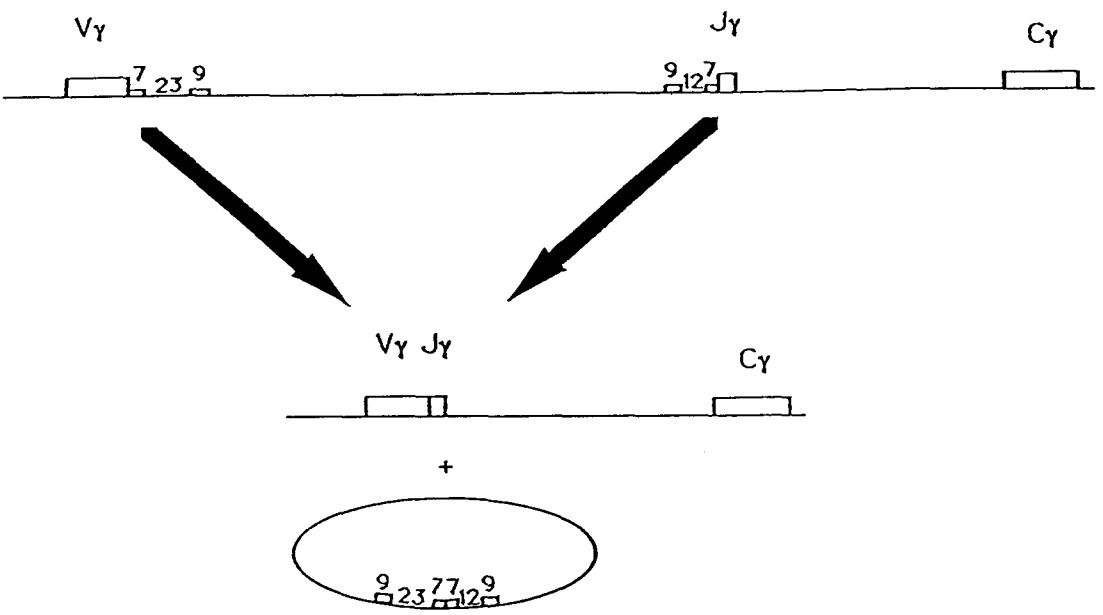


Figure 1. DNA rearrangement of a Vγ and a Jγ gene segment generates a Vγ gene. Top: A DNA fragment with a Vγ, a Jγ and a Cγ gene segment in germline configuration. Bottom: the products of the V to J rearrangement. The intervening DNA is excised and circularized. The recognition signals (7 $_{23}$ 9 and 9 $_{12}$ 7) are indicated.

MHC molecules.

The TCR and Ig differ in the way in which they recognize their antigens. Igs recognize the antigen directly, usually when the binding site (the epitope) is part of a larger structure. The TCR on the other hand can only recognize antigen in association with MHC (major histocompatibility complex) molecules. Antigen presenting cells (APCs) process the antigen and present proteolytic peptides on their MHC molecules to the extracellular environment. The phenomenon that T cells can only recognize antigen in the context of MHC molecules is termed MHC restriction.

There are two general types of MHC molecules: class I and class II. Class I molecules are expressed on the cell surface in association with ß2-microglobulin. The structure of a human class I molecule, HLA-A2, has been determined (Bjorkman *et al.*, 1987). It shows that the N-terminal part of the molecule forms a cleft between 2 alpha helices. This is where the peptide is bound and thus presented to T cells. It is not clear if all class I molecules can present peptide, even though the structure of most class I molecules is probably very similar. The classical class I molecules (K, D and L in the mouse) can present antigens and are expressed on virtually all cell types of the body.

There is also a large group of class I molecules for which no function has been found. These so-called class Ib

molecules (in the mouse also referred to as the TL and Qa antigens) are expressed on the cell surface of certain cells, also in association with ß2-microglobulin. Whereas the classical class I molecules are highly polymorphic, the class Ib molecules are far less polymorphic.

Class II molecules are heterodimers composed of alpha and beta chains. They present antigens like the classical class I molecules, but their tissue distribution is much more restricted than that of the classical class I molecules. While classical class I molecules present antigens to CD8 expressing cells (usually Tc), class II molecules present antigens mainly to Th cells, which leads to secretion of growth- and differentiation factors.

αß and γδ T cells.

Immunoprecipitation experiments showed that the TCR consists of 2 chains, which are noncovalently associated with a protein complex termed CD3 on the surface of T cells.

Most T cells in the body express the αß TCR. Their functions have been well characterized. They recognize specific antigen fragments when these are presented by MHC molecules on APC. As a population, the αß cells appear to be able to specifically recognize an enormous diversity of peptide antigens.

Like the αß cells, T cells expressing the γδ TCR have been

shown to be cytolytic and secrete factors as well. However, their functional specificity is not clear. There are specific γδ cell subsets, some of which show specific homing to mucosal and epithelial sites, some have invariant TCR variable domains, others shown antigen recognition without apparent MHC restriction and some subsets recognize only a single related group of antigens (Kyes and Hayday, 1990; Allison and Havran, 1991).

Table I. Nomenclature of Vγ genes, as proposed by Garman *et al.* (1986) and Raulet (1989) and Maeda *et al.* (1987).

Garman/Raulet	Maeda
V1.1	V1
V1.2	V2
V1.3	V3
V2	V4
V3	V5
V4	V6
V5	V7

Below is a review of the development of γδ T cells, their tissue localization and the antigens they may recognize. Next, the genes of the murine γ locus will be discussed. Table I shows the names of the 7 Vγ genes and the 2 different nomenclatures that are in use. For the purpose of

this discussion the nomenclature established by Garman *et al.* (1986) and Raulet (1989) will be used.

Ontogeny of γδ cells.

The first TCR γ RNA is detectable in the liver and gut (Vγ5), at day 11 of fetal development, before colonization of the thymus. In the fetal thymus, γ RNA is detectable on day 13, the first day investigated (Carding *et al.*, 1990). TCR ß RNA is first detectable on day 15 and α RNA on day 16 (Haars *et al.* 1986).

Serological analysis shows a particular time course of development. Cells expressing the γδ TCR appear in the thymus before day 14 of gestation and increase in number up to day 20. Cells expressing αß TCRs start to appear at day 16 and become the dominant population around day 19 (Havran and Allison, 1988; Itohara *et al.*, 1989).

Distinct waves of γδ cells arise in the thymus.

Most thymocytes in the day 14 and 15 fetal thymus express a $V\gamma3^+$ γδ TCR on the cell surface, a population which rapidly declines and is scarcely detectable by day 18 (see figure 2). This early population uses exclusively the V3J1Cγ1 and V1D2J2Cδ chains in the TCR (Havran and Allison, 1988; Ito *et al*, 1989; Lafaille *et al*, 1989). The Vγ3 cells are progressively replaced by cells expressing V4J1Cγ1 and

V1DJδ2 (Ito *et al.*, 1989; Allison and Havran, 1991). These cells in turn are replaced by Vγ2 expressing cells and finally Vγ5 cells appear. Serologically, Cγ4 containing TCR were detected in the fetal thymus as early as day 16. Earlier time points were not investigated, nor was the V which it contained determined (Houlden *et al.*, 1988). Interestingly, the order of appearance of the Vγ2-Vγ5$^+$ cells in ontogeny parallels the order of the V gene segments on the chromosome (see figure 2).

The early Vγ3 and 4 expressing thymocytes have an unusual feature: their in frame V-Jγ and V-D-Jδ junctions are almost invariant. There is limited exonuclease activity evident in these canonical junctions, and no N region nucleotides are observed (Lafaille *et al.*, 1989). The V1D2J2Cδ genes and V-Jγ1 junctions in the Vγ3$^+$ and Vγ4$^+$ cells are identical. Since Vγ3$^+$ and Vγ4$^+$ cells with these same canonical junctions were found in specific locations in the periphery, it has been suggested that the fetal canonical cells are their precursors (Lafaille *et al.*, 1989; Asarnow *et al.*, 1989; see below). Indeed, early fetal Vγ3$^+$ thymocytes have been shown to be the only cells which can give rise to the canonical Vγ3 T cell population in the periphery (see below, Havran and Allison, 1990). The precursor-product relationship of the fetal Vγ4 population has not been investigated.

The sequential appearance of γδ T cells expressing different TCR Vγ genes in the thymus raised the question whether this is due to preprogrammed precursor cells, or due to instruction by the thymic environment.

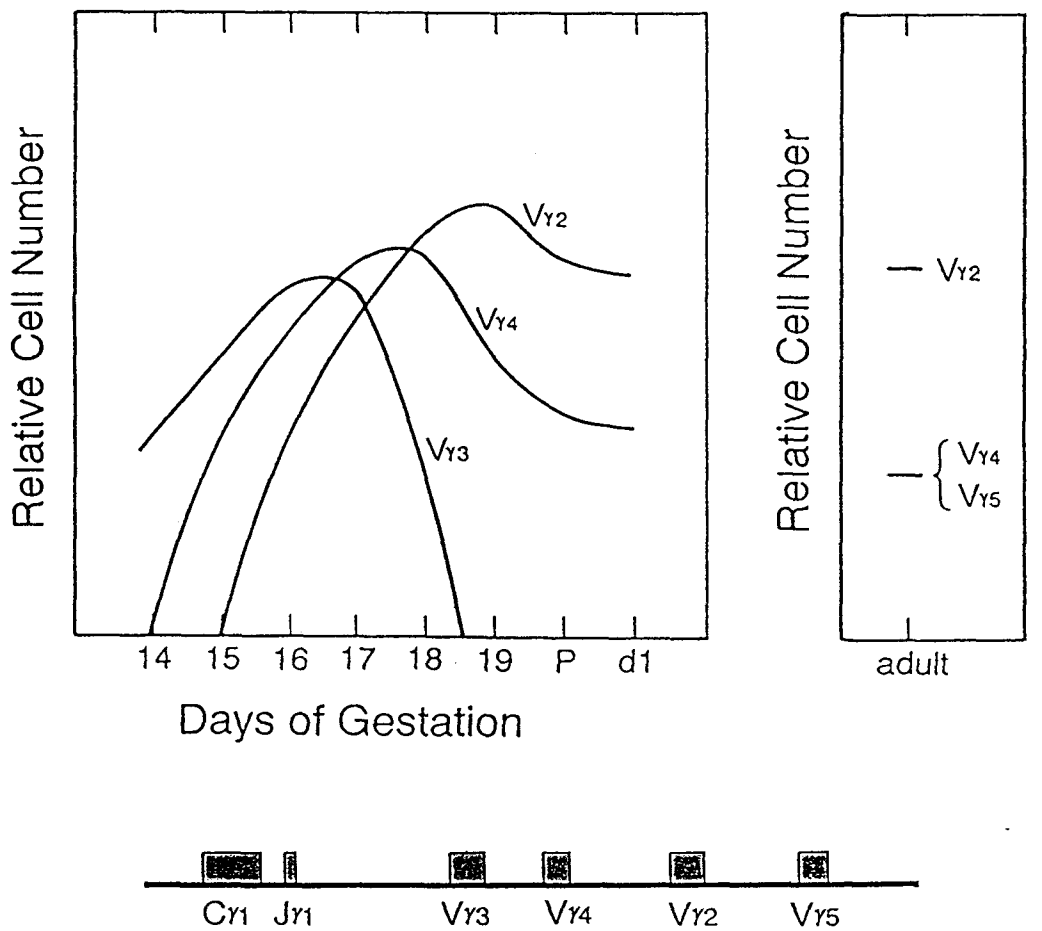


Figure 2. The order of expression of the Vγ genes in the fetal thymus parallels the gene order in the γ locus.

Experiments showed that only fetal stem cells give rise to significant numbers of $V\gamma3^+$ cells. Fetal and adult stem cells can give rise to $V\gamma4^+$ cells, but only the RNA from fetal derived cells encoded invariant $VJ\gamma$ junctions identical to those of normal fetal $V\gamma4^+$ cells (Ogimoto *et al.*, 1990; Ikuta *et al.*, 1990; Ikuta and Weissman, 1991). These results suggest that there may be a developmental clock in the stem cells which determines when to close or activate certain genes (e.g., $V\gamma3$ and N region insertional machinery), or alternatively that there are different types of stem cells, each with its own developmental program (Ikuta *et al.*, 1990; Ikuta and Weissman, 1991).

Extrathymic γδ T cell maturation.

Nude mice lack a thymus, and are virtually devoid of αß T cells. T cells expressing γδ TCRs on the other hand can be found in the gut, the spleen and lymph nodes of nude mice, albeit in decreased numbers (Pardoll *et al*, 1988; Yoshikai *et al.*, 1986; Bandeira *et al.*, 1991).

Intestinal intraepithelial lymphocytes (i-IEL) with functionally rearranged $V\gamma1.2$, 2 and 5 develop in athymic mice as well (Whetsell *et al.*, 1991). Hence, it has been suggested that the gut epithelium attracts progenitor cells and induces maturation, including rearrangement of the TCR genes (Mosley *et al.*, 1990; Guy-Grand *et al.*, 1991).

There is also evidence that γδ expressing resident pulmonary lymphocytes (RPL) can develop extrathymically, as they too are present in nude mice (Sim and Augustin, 1990; 1991). Mature $V\gamma3^+$ cells reside in the skin and are absent in nude mice, indicating that they are thymus dependent (Havran and Allison, 1990). However, Vγ3 and 4 may be able to rearrange in nude mice (unpublished results in Kyes and Hayday, 1990).

Thus, it appears that most if not all Vγ genes can rearrange extrathymically and at least some γδ sublineages can mature outside the thymus as well.

The presence of some γδ populations in nude mice raises the question whether there are 2 maturation pathways for (or 2 sublineages of) γδ cells, one occurring in the thymus and one outside the thymus (Mosley *et al.*, 1990; Guy-Grand *et al.*, 1991) or whether γδ cells can develop outside the thymus, but that this process is much more efficient inside the thymus (Pardoll *et al.*, 1988).

Tissue distribution.

Maybe the most remarkable feature of γδ cells is that there is a specific tissue distribution for some of the γδ subsets. Table II and figure 3 summarize the phenotype of γδ cells in specific tissues.

The skin.

The dendritic epithelial cells (DECs) in the skin are thy-1$^+$, $\gamma\delta^+$ cells which are interspersed between the keratinocytes in the outer layer of the skin (Elbe *et al.*, 1989). The $\gamma\delta$ TCR is identical in all DECs: these cells all use exclusively V3J1Cγ1 and V1D2J2Cδ. Even the V(D)J

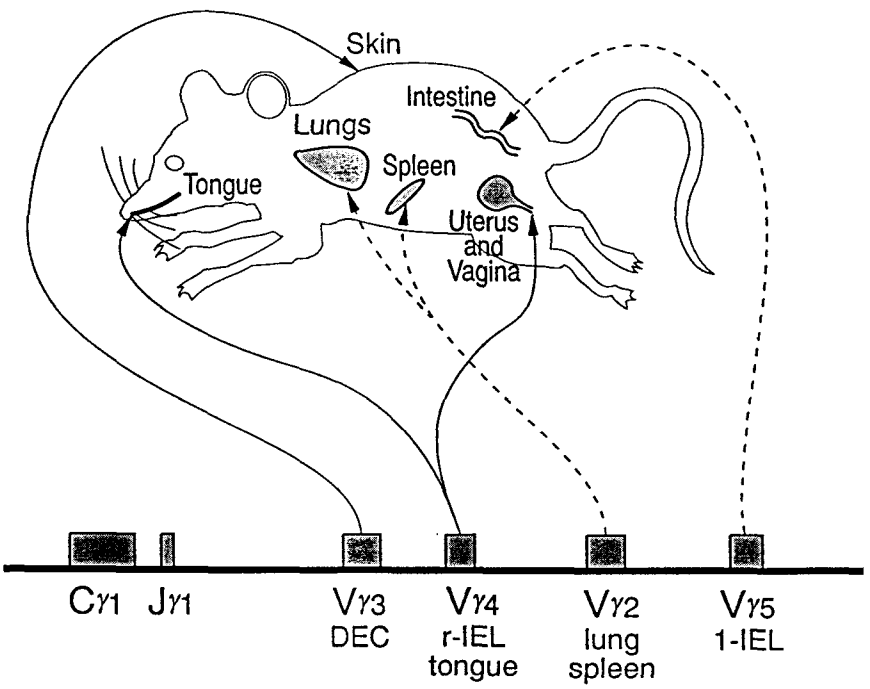


Figure 3. Schematic view of the tissue distribution of Vγ2-5$^+$ $\gamma\delta$ subsets in adult mice. Solid lines indicate near exclusive homing and broken lines indicate preferential localization (adapted from Cheng *et al.*, 1991).

junctions of the 2 TCR chains are nearly invariant in the DECs (see table III; Havran *et al.*, 1989; Asarnow *et al.*, 1988, 1989).

As a population the DEC are $V\gamma3^+$. However, three DEC cell lines have been reported which express $C\gamma4$ and/or $C\gamma2$ proteins on their surface, with $V\delta1$ (Koning *et al.*, 1988; McConnel *et al.*, 1989). The significance of this is not known.

Thymic precursors. There is only one other discrete T cell population which expresses $V\gamma3$ on its surface, namely the earliest wave of $\gamma\delta$ cells in the fetal thymus (Havran *et al*, 1988). The 'canonical' V(D)J junctions of the fetal γ and δ TCR rearrangements (see above) are identical to those of the DECs (table III; Asarnow *et al.*, 1988; 1989). Indeed, functional DECs only derive from early fetal cells (Havran and Allison, 1990; Payer *et al.*, 1991).

Vagina, uterus, tongue.

Vγ gene. Approximately half of the T cells in the vagina, uterus and tongue carry $\gamma\delta$ TCRs on their surface. The most predominant $V\gamma$ on these cells is $V\gamma4$ (table II, figure 3). All the analyzed in frame γ junctions were identical to those of the fetal canonical V4-Jγ1 sequence (Itohara *et al*, 1990a; Nandi and Allison, 1991).

Table II. Tissue distribution of γδ subsets in adult mice[a].

tissue	major TCR		diversity[b]
skin (DEC)	V3J1Cγ1	V1D2J2Cδ	limited
r-IEL	V4J1Cγ1	V1D2J2Cδ	limited
lung (RPL)	V2J1Cγ1	Vδ6/5	++
intestines	V5J1Cγ1	V4(D1)D2Jδ1	+++
(i-IEL)		V6(D1)D2Jδ1	
lactating mammary gland	V2/V3-J1Cγ1	Vδ4	+++
thymus	V2/V4-J1Cγ1	Vδ5/6	+++
	V1.1J4Vγ4		
	V1.2J2Cγ2		
spleen	V1.1J4Cγ4	Vδ5/6	+++
	V2J2Cγ1	Vδ4/2	
	V1.2J2Cγ2		

[a]Abbreviations: r-IEL, reproductive IEL; i-IEL, intestinal IEL; RPL, resident pulmonary lymphocytes.

[b]Diversity at the V-J junction of the rearranged γ gene segments. limited: the V-J junctions are invariant as described in the text; +++: the junctions are very diverse, ++: the junctions are very diverse with the exception of one group (see text).

Adapted from Cheng *et al.* (1991).

Vδ gene. The only Vδ gene segment found in γδ cells of these tissues is Vδ1 (table II). In the majority of cases, this Vδ is rearranged to Dδ2 and Jδ2, and has a junction which is identical to that of the fetal canonical sequence and the

DEC V1DJδ junction (see above and table III; Nandi and Allison, 1991; Itohara *et al.*, 1990b). Thus, it is possible that the fetal Vγ4 population contains the precursors for these peripheral $V\gamma4^+$ populations, analogous to the DEC lineage. The absence of γδ cells from the vagina of nude mice suggests that the precursors need the thymus for maturation (Nandi and Allison, 1991).

Intestine.

The proportion of i-IEL which express a γδ TCR appears to vary widely (30-80%). Generally, about half of the i-IEL are γδ T cells, the rest are αß T cells. The γδ cells are interspersed between the epithelial cells on the villi in the intestines (Bandeira *et al.*, 1991).

Vγ genes. Vγ5 is the predominant γ gene in the i-IEL population, but Vγ2 and 4 RNA are also present (table II, Takagaki *et al.*, 1989a; Kyes *et al.*, 1989; Bonneville *et al.*, 1988). PCR amplification of the V-Jγ junctions showed that a large proportion of junctions with Vγ2 and 4 are out of frame, while most V5Jγ1 junctions are in frame and are

Table III. Compilation of frequently used or unusual V-Jγ junctions.

tissue	TCR V gene junctions
DEC, early fetal thymocytes	
γ: EATYY	C A C W D S S GF TGT GCC TGC TGG GAT AGC TCA Vγ3 / Jγ1
δ:	C G S D I G G S S W TGT GGG TCA GAT ATC GGA GGG A G C TCC TGG Vδ1 / Dδ2 /N/ Jδ2
r-IEL, fetal thymocytes	
γ: TGTYY	C A C W D S S GF TGT GCA TGC TGG GAT AGC TCA Vγ4 / Jγ1
δ: same as in DEC	
RPL in lung	
γ: GxYS	C S Y G x Y S G TGT TCC TAC GGC NNN T AT AGC TCA Vγ2 / N / Jγ1
δ: BID	C A S G Y I G G I R A T D TGT GCC TCG GGG TAT ATC GGA GGG GAT CGA G CT ACC GAC Vδ5 / Dδ2 / Jδ1

From Lafaille *et al.* (1989), Asarnow *et al.* (1988), Asarnow *et al.* (1989), Itohara *et al.* (1990a, 1990b), Nandi and Allison, (1991), Allison and Havran (1991), Sim and Augustin (1990, 1991).

reported as quite diverse (Asarnow *et al.*, 1989; Kyes *et al.*, 1989; Takagaki *et al.*, 1989a). Only one report stresses the low variability in the junctions (Whetsell *et al.*, 1991). However, close inspection of the other reported Vγ5 V-J junctions shows that within each report half or more of the junctional sequences are identical to one another, or differ by 1 amino acid.

Vδ genes. The i-IEL express predominantly Vδ6 and 4, with a smaller number of cells expressing Vδ5, 7 and 1. The junctions of these TCR genes show considerable diversity at the nucleotide and amino acid level (Takagaki *et al.*, 1989a; Kyes *et al.*, 1989; Asarnow *et al.*, 1989). Thus, there is no restricted Vδ gene segment usage.

There appears to be selection or preferential homing of Vδ4 expressing γδ i-IEL in I-E^k mice, as these mice have more $V\delta4^+$ i-IEL than non I-E^k mice. This selection occurs extrathymically (Lefrancois *et al.*, 1990). Since intestinal epithelium cells express class II, they may be involved in this selection process (Lefrancois *et al.*, 1991; Guy-Grand and Vassalli, 1986).

In summary, other than a predominance of Vγ5, there appears to be no restricted diversity of i-IEL γδ TCR, in contrast to the γδ TCRs in skin and vut IELs.

Lung.

The T cells in the lung are referred to as resident pulmonary lymphocytes (RPL). The γδ T cells constitute approximately 28% of all $CD3^+$ T cells in the lung (Augustin *et al.*, 1989).

Vγ genes. The γδ cells in the lung predominantly use Vγ2 (table II). In BALB/c mice, most (90%) of these have a common feature: they show a junction which encodes the amino acid sequence GxYS, where x can be 'any' amino acid, and G and S are encoded by germline V and J sequences, resp. (table III). RPLs of C57BL/6 mice have this sequence in only about 40% of the $V\gamma2^+$ cells. Extrathymic selection of Vγ2 GxYS junctions is suggested by the fact that the functional Vγ2 rearrangements in the thymuses of these mice show only limited numbers of GxYS junctions and by the fact that nude mice also have a high proportion of GxYS junctions in the $V\gamma2^+$ RPL (Sim and Augustin, 1991).

Vδ genes. The repertoire of Vδ genes expressed in RPLs is highly diverse (table II). The most abundant δ RNA is Vδ6, followed by Vδ5, then Vδ7 and Vδ4. RNA for Vδ3 and 2 is very low (Sim and Augustin, 1990). They are all rearranged to Jδ1. In adult BALB/c these junctions are highly diverse (N nucleotides, use of 1 or both Ds), except in the Vδ5 junctions.

Over 80% of the characterized in frame V5Jδ1 junctions use

only Dδ2, and 2/3 of these are invariant, with no N nucleotides and no loss of germline sequences. Thus, over 60% of the productive Vδ5 rearrangements have identical junctions in BALB/c RPL. These Vδ5 junctions have been named BALB/c invariant delta (BID), since they have not been detected in C57BL/6 mice. It appears to be dominantly inherited, since F1(BALB/cXC57BL/6) also show this feature in their rearranged Vδ5 RPL junctions. The BID sequences are also present among RPLs in nude Balb/c mice, suggesting that their predominance is caused by positive selection (expansion) in the periphery. Since BID is also present in lymph nodes, the selection may occur at sites outside the lungs. BID is expressed on a small fraction of the RPLs and GxYS on a majority, making it is unlikely they are selected for together (Sim and Augustin, 1990).

Spleen and lymph nodes.

The γδ cells in the spleen, lymph nodes and blood are all Thy-1$^+$, nearly all CD8$^-$, and make up 3-5% of all CD3$^+$ cells in the spleen and lymph nodes (Itohara *et al.*, 1989; Cron *et al.*, 1988). The first T cells appear to enter the spleen after birth.

Vγ chains. Adult splenic γδ cells predominantly express V1.1Cγ4, V2J1Cγ and V1.2Cγ2 (Cron *et al.*, 1990; Ezquerra *et al.*, 1990). Nude mice accumulate only small numbers of αß

and γδ cells in the spleen, with the γδ cells expressing Cγ1, V1.2-Cγ2 and Cγ4. This indicates that the splenic γδ cells are mainly thymus dependent (Pardoll *et al.*, 1988; Cron *et al.*, 1990).

δ chains. The splenic delta chains show extensive junctional diversity (Lacy *et al.*, 1988). Vδ5 appears to be the predominant Vδ used in adult splenocytes, with Vδ2, 4, 6 and Vα10 present as minor species (Ezquerra *et al.*, 1990).

Thymus.

Vγ chains. The γδ cells develop in the thymus in an ordered pattern (see above). The first to appear are Vγ3/Vδ1 expressing cells, followed by Vγ4/Vδ1 and then Vγ2 and Vγ5 cells (see figure 2). V1.1J4Cγ4 is also expressed in the fetal thymus, as early as day 16 (Houlden *et al.*, 1988). Adult thymocytes express mainly Vγ2, and some Vγ5 and 4 (Takagaki *et al.*, 1989b; Korman *et al*, 1988) Rearrangements of Vγ3 and 4 were not detected in adult thymus DNA, whereas rearrangements of Vγ2 and one or more V1 members appear to be quite frequent (Garman *et al.*, 1986). A prominent difference between fetal and adult junctions is the presence of considerable variability in the adult junctions, as compared to fetal junctions, both in the γ and δ junctions (Lafaille *et al.*, 1989).

Vδ chains. Vδ5 and 6 appear to be the predominant Vδs that

are expressed in the adult thymus (Takagaki *et al.*, 1989b; Elliott *et al.*, 1988). One of the Vδ7 genes has been reported to be rearranged in a large number of thymocytes also (Korman *et al.*, 1988). Early fetal thymocytes use Dδ2 and Jδ2 frequently (see above), but adult thymocytes generally use both D's and Jδ1 in their Vδ rearrangements (Elliott *et al.*, 1988).

Homing of γδ cells to epithelial sites.

As described, certain γδ cell populations reside at particular anatomical sites in the body. Whether the localization was due to the particular γδ receptor that they displayed was investigated in transgenic mice. Vγ2/Vδ1 transgenic mice express this transgene encoded TCR on all γδ cells including on skin and i-IEL γδ cells. This shows that homing of γδ cells is not determined by the nature of the γδ TCR. The nature of the homing determinant is not known (Bonneville *et al.*, 1990; Barrett *et al.*, 1992).

γδ cells in other vertebrates.

The pattern of γδ cell expression in mammals is diverse. In the mouse, the γδ cells make up 5-10% of the peripheral blood lymphocytes and some subsets show tissue specific localization.

Human. The γδ cell content of human peripheral blood lymphocytes (PBL) is variable but generally low, around 5%. The spleen and lymph nodes contain similar proportions of γδ cells. A substantial, but variable fraction of these the γδ PBL express CD8 (Groh *et al*, 1989). Most of the γδ PBL express V9J1Cγ1 and Vδ2, even though these cells are rare in the thymus. Peripheral expansion by selection has been suggested (Parker *et al.*, 1990).

In human, a minority of the i-IEL are γδ cells, but this population seems to be increased moderately over that in the blood, indicating a somewhat preferential localization (Ullrich *et al.*, 1990). Nearly all γδ i-IEL are located in the epithelial cell layer, rather than the lamina propria. There is no evidence for preferential usage of Vγ genes in the human i-IEL, but here does seem to be a preferential use of Vδ1. The V(D)J junctions of the γ and δ genes show considerable diversity however (Ullrich *et al.*, 1990; Jarry *et al.*, 1990; Deusch *et al.*, 1991). The skin has a low number of γδ cells which are outnumbered substantially by αß cells, a situation unlike that in the mouse.

These results show no predominant expression of γδ cells at specific locations in the human body, other than in the gut epithelium (Groh *et al*, 1989; Ullrich *et al.*, 1990).

Ruminants. The γδ cells in sheep and cow constitute a large

part of the circulating T cells (60%), unlike in the mouse and human. In the ruminants, only a small minority of the T cells in the peripheral lymphoid tissues are γδ cells. They reside in large numbers in epithelia such as the gut and tongue and in cattle also in the skin (Hein and Mackay, 1991). Thus, the epithelial localization observed in the mouse is also present in these mammals. The TCR V genes that are used in these populations are not known.

Birds. Chickens also contain large numbers of γδ cells. The γδ cells in the blood and spleen are 20-50% of the total lymphocyte population. They are also prominent in the epithelium of the intestines, but not in the skin (Bucy *et al.*, 1988; Cooper *et al.*, 1989).

Function.

The physiological function of the immune system is to protect the body from foreign pathogens and neoplastic cells. How the γδ cells fit in this system is not entirely clear. As described above, much more is known about their differentiation and tissue localization, but little about what they do. The αß cells have been well studied. When an animal is injected with an antigen, αß cells which are specific for that antigen can be readily isolated. This is not the case with γδ cells (see below).

The possible function of γδ cells has been the topic of speculation. The preferential localization of γδ cells in various epithelia has suggested that the γδ cells are involved in surveillance of the bodies surfaces, or that they are the first line of defense against infections. The monomorphic nature of the TCR of several of these subsets has suggested that they must recognize an equally monomorphic ligand, probably a self protein, which may be presented in response to infection (Janeway, 1989).

In order to gain insight into the function of the γδ cells, much effort has been put in defining the structures these cells recognize. As described below, some of the ligands are possibly microbial antigens and (altered) self molecules, to which the γδ cells respond by secretion of lymphokines and cytotoxicity. Class Ib molecules have been implicated as restriction elements for 1 subset.

1. *MHC products.*

The αß T cells exhibit class I or class II MHC restricted antigen recognition, that is, they recognize their peptide antigens as part of a peptide-MHC molecule complex. Hence, it comes as no surprise that many αß cells can recognize class I or class II MHC molecules from H-2 disparate mouse strains, presumably because they mimic a 'self MHC + peptide' structure. In contrast, γδ cells do not show this

strong bias to recognition of allogeneic MHC class I and class II molecules. If anything, the isolated γδ cells that do bind MHC molecules recognize class Ib (TL and Qa) molecules as often as they do class I and class II molecules.

Class I and class II reactivity.

Several cell lines (lines or hybridomas) expressing γδ TCRs reactive with class I or class II molecules have been isolated. Table IV lists the mouse clones and table V the human clones which recognize well defined target molecules (with or without antigen).

Mouse clones LBK5 and LBK1 both recognize class II molecules. They use the same γ and δ genes, but the junctions are different.

Several of the clones were isolated from nude mice (LBK5 and -1, unnamed H-2D reactive clone, G8). The reactivity of clone G8 could be inhibited by anti CD8 antibodies (Bluestone *et al.*, 1991).

Hybridoma KN6 recognizes a TL molecule, which is expressed on many cell types. The hybridoma was derived from C57BL/6 thymocytes, without selection (Bonneville *et al.*, 1989; Ito *et al.*, 1990). The V2Jγ1 junctions of KN6 and clone G8 are different and the 2 cells use different Vδ genes (Bluestone *et al.*, 1991).

One mouse γδ hybridoma is known for which the restricting element and the presented epitope are identified. This cell

Table IV. Mouse MHC reactive γδ T cell lines and hybridomas.

name	reactivity	Vγ/Vδ	isolation[a)]	reference
LKB5	I-$E^{k,b,s}$	V1.2Jγ2	MLC	2
		V5D2Jδ1		
LKB1	I-A^{d}	V1.2Jγ1	MLC	3
		V5D2Jδ1		
-	H-$2D^{k}$	Cγ4	MLC	4
G8	TL^{k}	V2Jγ1	MLC	1, 4
		Vα11D1D2Jδ1		
KN6	TL-$27/22^{b}$	V2Jγ1	random	5, 6
		V5DJ1Cδ		
DGT3	Qa-1+GT	Cγ1	GT	7, 8
CGT3				

a) The clones were isolated after mixed lymphocyte culture (MLC) or immunization (GT). Hybridoma KN6 was made from a fusion with normal $\gamma\delta^{+}$ thymocytes.

References: (1) Matis *et al.* (1987), (2) Matis *et al.*, (1989), (3) Rellahan *et al.* (1991) (4) Bluestone *et al.* (1988), (5) Bonneville *et al.* (1989), (6) Ito *et al.* (1990), (7) Vidovic *et al.* (1989), (8) Vidovic and Dembic (1991).

can respond to a copolymer of poly (Glu^{50} Tyr^{50}) (GT), unless antibodies to Qa-1^b are present (Vidovic *et al.*, 1989). Furthermore, the hybridomas were capable of providing specific (Qa/GT) B cell help, i.e., they can stimulate antibody production (Vidovic and Dembic, 1991).In human, several target structures for γδ cells have been described, including one example of a restricted response to a known antigen (see table V).

The specificity of some clones was characterized in great detail, including after expression of the target structure in mouse cells (clones LM12, ES-204 and ES-433). The reactivity of other clones was only determined by antibody blocking of the target structure (e.g., clone RNG-135).

CD1c is a class Ib molecule, which is associated with ß2m on the cell surface. Two independent reports show reactivity of isolated γδ cell lines to CD1c. Anti CD1c antibodies blocked the response (Porcelli *et al.*, 1989; Faure *et al.*, 1990). CD1c is expressed on a small number of cells in human, including subsets of T and B cells and Langerhans cells in the skin.

TCT.1/Blast-1/CD48 is not an MHC molecule, but is a member of the immunoglobulin supergene family. It shares homology to CD4, class IIα and V-kappa proteins and is anchored to membrane lipids (Staunton and Thorley-Lawson, 1987). γδ T cells specific for TCT.1/Blast-1/CD48 have been isolated

from 2 individuals. The molecule is widely expressed on haematopoietic cells, but not on various other cell types (Mami-Chouaib *et al.*, 1991; Del Porto *et al.*, 1991).

Table V. Specificity of human γδ T cell lines and hybridomas.

name	reactivity	isolation	reference
LM12	HLA-A24	MLC	1
ES-204	HLA-A2	MLC	2
ES-443	HLA-A2	MLC	2
RNG-135	HLA-DQ	MLC	3
N2A11	HLA-DR7	random	8
DF	HLA-DR4 + TT	immunization	4
J2B7	CD1c	random	5
IDP2	CD1c	patient	6
E,G	TCT.1/Blast-1/CD48	MLC	7

---

a) The clones were isolated after mixed lymphocyte culture (MLC), immunization, from an immunodeficiency patient, or from randomly generated $\gamma\delta^+$ PBL clones.

References: (1) Ciccone *et al.* (1989), (2) Spits *et al.* (1990) (3) Bosnes *et al.* (1990), (4) Kozbor *et al.* (1989), (5) Faure *et al.* (1990), (6) Porcelli *et al.* (1989), (7) Mami-Chouaib *et al.* (1991), (8) Jitsukawa *et al.* (1988).

Despite efforts to isolate alloreactive γδ cells, few such clones have been isolated. It appears much easier to isolate allospecific αß cells than allospecific γδ cells (O'Brien *et al.*, 1989). Since some alloreactive γδ cells have been isolated, it may be that classical and nonclassical MHC molecules can function as restricting molecules for at least certain classes of γδ cells. Yet the low frequency of MHC reactive γδ (relative to αß) cells may be an indication that (1) MHC restricted γδ cells are a minority in the γδ T cell population, and/or

(2) most MHC reactive γδ cells are sequestered in epithelia and not present in blood or peripheral lymphoid tissues (the cell lines and hybridoma in tables IV and V were not derived from epithelial γδ populations), and/or

(3) the structure and/or selection of the TCR Vγ and Vδ genes predisposes the γδ cells against self recognition (is a large part of the γδ TCR binding affinity due to interaction with the ligand, making the TCR more 'ligand specific' than 'MHC specific'?).

It should be mentioned that using MLR to generate alloreactive γδ cells (which is how most clones in tables IV and V were isolated) may predispose the experiment to generate alloclass I and alloclass II γδ cells, rather than alloclass Ib γδ cells. Only when there are differences

between the surface molecules expressed by the 2 cell types in the MLC (stimulator cells and responder cells) is activation of responder T cells expected. The class I and class II molecules are much more polymorphic than the class Ib molecules. They are therefore much more likely to be different between the 2 celltypes in an MLC with cells from 2 random individuals or mouse strains than the class Ib molecules. In addition, in an MLC, reactive γδ cells can only be generated against the surface molecules present on the stimulating cells. Thus, γδ cells directed against, for instance, a gut specific surface molecule can not be generated if the stimulator cells are spleen cells.

If certain class Ib molecules are not expressed on stimulator cells, then the MLR would not allow the isolation of γδ cells specific for those class Ib molecules.
For instance, class Ib molecules which are expressed on gut epithelial cells may not be expressed on spleen or blood stimulator cells. Thus, this MLC cannot generate γδ clones capable of recognizing this gut antigen.

In summary, finding γδ cells reactive with classical class I or class II molecules is not surprising due to the selection for them by MLR. To find approximately equal numbers of class Ib reactive γδ cells is interesting and may

point to a function of the class Ib molecules in γδ reactivity. There are indeed data supporting the use of TL molecules as restriction elements for one γδ subpopulation. The response of isolated i-IEL to isolated gut epithelial cells can be inhibited by antibodies to Tla (Eghtesady and Kronenberg, 1992).

The CD8 molecules interacts with the α3 domain of classical and nonclassical class I molecules (Salter *et al.*, 1990; Teitel *et al.*, 1991). Hence, it is no surprise that part of the γδ i-IEL express CD8 on the cell surface. The DECs in the skin and most γδ cells in other anatomical locations (spleen, lymph nodes, thymus), do not express CD8. But when maintained in vitro, the percentage of CD8 expressing cells increases. This raises the question whether only specific γδ populations use class I/Ib molecules as restriction elements (such as the i-IEL), or whether many or most γδ cells use these molecules as restiction elements, but only after activation induces CD8. Alternatively, γδ T cell populations may exist which use class I/Ib restriction elements, but which do not require (nor express) CD8 for this interaction. The answer to this question is not clear.

Class Ib molecule expression in the body is variable. RNA from the T3, T9 and T21 genes is present in very few BALB/c mouse tissues, but it is abundant in small intestinal

epithelial cells. RNA from the T10 and/or T22 and M2 genes is also present in mouse intestinal epithelium. In situ antibody staining showed indeed that T3 and possibly T18 are present on the epithelial cell surface (Wu *et al.*, 1991; Hershberg *et al.*, 1990). Also expressed on the mouse gut epithelium is CD1 (Bleicher *et al.*, 1990). Two human CD1c reactive γδ T cells have been isolated (table V). The T22 and T23 genes are expressed on a wide variety of tissues (Ito *et al.*, 1990).

Thus, various class Ib molecules are expressed at various places in the body (see also Strominger, 1989; Stroynowski, 1990).

2. Viral and bacterial antigens.

Only one example of a γδ clone which responds to a known viral antigen is known. A γδ T cell directed against Herpes Simplex Virus surface glycoprotein I was raised by virus infection and subsequent coculture of lymphnode cells with cells expressing this molecule. This response was not restricted by class I or class II molecules. Vγ1.2 was used in this clone (Johnson *et al.*, 1992).

Bacterial antigens in mice. O'Brien *et al.* (1989) showed that a substantial portion of γδ expressing hybridomas derived from unselected thymocytes from newborn mice

spontaneously produced IL-2, indicating that they were active in culture without further stimulation. This activity was inhibited by antiCD3 antibodies, suggesting that the cells recognize a cell- or medium-derived structure with their TCR. All of these spontaneously active clones could be further stimulated by purified protein derivative (PPD) from Mycobacterium tuberculosis. And nearly half of these cells could be stimulated by a 65 kD heat shock protein (hsp-65, a component of PPD) from a related Mycobacterium, although this response was weaker than the PPD response. This indicated that an evolutionary highly conserved molecule (hsp-65) was recognized by these γδ hybridomas (O'Brien *et al.*, 1989; O'Brien and Born, 1991; O'Brien *et al.*, 1991). An hsp-65 peptide which stimulated all tested PPD reactive hybridomas strongly, showed homology to the corresponding sequence of the hsp-65 homologs in mouse/human and yeast. Synthetic peptides of these yeast and mammalian hsps indeed stimulated 2 tested clones, although the response was weaker than with mycobacterial hsp peptide. This result led to the speculation that PPD reactive γδ cells might crossreact with autologous hsp, explaining the frequent spontaneous stimulation of γδ clones in culture (Born *et al.*, 1990). Interestingly, all of the analyzed PPD reactive clones expressed functional V1.1J4Cγ4, mostly (25/28) together with one of 2 Vδ6 gene family members. The junctions of the

rearranged genes showed a somewhat limited diversity. Most, but not all V1.1J4Cγ4 expressing hybridomas were PPD reactive (Happ *et al.*, 1989).

The response of γδ cells to mycobacterial antigens is not limited to these thymocyte hybridomas. A subset of hybridomas from adult γδ spleen cells also secreted IL-2 without stimulation. These clones all expressed V1.1J4Cγ4, like the thymocyte hybridomas, but the junctions were more diverse than in the newborn thymocyte hybridomas (O'Brien and Born, 1991; O'Brien *et al.*, 1991).

Several other lines of evidence show that γδ cells respond to mycobacterial antigens. Immunizing mice with mycobacteria resulted in a large expansion of the lymph node γδ cells (Janis *et al.*, 1989) in concert with an expansion of αß cells. In a secondary response, the number of γδ cells did not go up, but the αß cells increased in number faster than in the primary response (Griffin *et al.*, 1991). Nude mice also contain PPD reactive γδ cells in the lymph node, which also expand after immunization (Yoshikai *et al.*, 1990). Aerosols containing PPD also increased the numbers of γδ cells, but this time in the lung (Augustin *et al.*, 1989). Thus, proliferation of γδ T cells appears to be a general result of exposure to mycobacterial antigens.

Bacterial antigens in human. Human γδ cells have also been shown to proliferate *in vitro* in response to mycobacterial antigens. All the proliferating cells express Vγ9/Vδ2. However, this response could be a superantigen response. The junctional diversity of the V genes was very large (indicative of a polyclonal expansion; Band *et al.*, 1991; Ohmen *et al.*, 1991) The nature of the antigen(s) is not clear. It could be the mycobacterial hsps (Fisch *et al.*, 1991). A low molecular weight component of a mycobacterial lysate has also been described which requires class II expression for its stimulation of a large proportion of $V\gamma9^+$ cells (Pfeffer *et al.*, 1992).

Other sources of human γδ superantigens have been described including a component on Daudi cells (probably an hsp; Sturm *et al.*, 1991; Fisch *et al.*, 1990) and Staphylococcal Enterotoxin A (SEA), which also requires class II in order to stimulate the $V\gamma9^+$ cells (Rust *et al.*, 1990).

Antigen specific responses involving mycobacterial lysates occur as well: several hsp-65 reactive clones have been isolated (Holoshitz *et al.*, 1989; Haregewoin *et al.*, 1989), even though this group may be a minority in the γδ subset which is reactive to mycobacterial antigens (Kabelitz *et al.*, 1990).

In summary, the response of human γδ cells to mycobacterial antigens is strong, but what the primary antigen is, has not

been shown conclusively. It could be the mycobacterial hsp, a superantigen of mycobacterial origin, or even a self component induced by the mycobacterial antigen. Whether the use of mouse Vγ1.1 in the response to PPD is akin to the human Vγ9 response to mycobacterial antigen is not clear.

3. Stress proteins.

There are a number of experiments that suggest that γδ cells may recognize stress proteins. In the previous section, spontaneously IL-2 secreting thymocyte hybridomas are described which do not need antigen presenting cells (APCs) for the spontaneous activity. Blocking the γδ TCR blocks the IL-2 release, indicating that the stimulation is through the TCR. Some of these clones respond to mycobacterial hsp-65 and also to a peptide derived from it, as well as to the homologous peptide of mouse hsp-63, although weaker.

Selective expansion of γδ cells occurs when isolated lymph node or pulmonary cells are subjected to heat shock and then cultured, without antigenic stimulation. The heat shock induced a number of hsps. The expansion was even stronger if the cells were first exposed in vivo to mycobacterial antigens, prior to the in vitro heat shock (Rajasekar *et al.*, 1990).

At least 2 of the epithelial γδ cell populations appear to

respond to stressed cells: DECs and i-IEL. Virtually all DECs express identical TCR, with identical junctions (see above). This observation already led to the speculation that these cells must recognize a very monomorphic ligand, probably not some unknown foreign antigen, but a self protein (Asarnow *et al.*, 1988).

Freshly isolated DECs respond to freshly isolated keratinocytes from the skin and to a keratinocyte cell line, but not to fibroblasts or spleen cells. DEC γδ clones were the only T cells that could respond to the keratinocyte cell line. The recognition was shown to be dependent on the TCR and not MHC restricted, as DECs from MHC disparate strains responded to keratinocytes of all tested strains and antiMHC and antiQa antibodies did not inhibit the response. The DECs can also respond to fibroblasts that have been incubated with tryptic digests of keratinocytes.

The authors concluded that the γδ DECs recognize a self antigen which is present or produced in isolated keratinocytes. Since the response could be enhanced by heat shocking the keratinocytes, the ligand that is recognized is probably derived from a stress protein, or induced by it. The stress molecule is probably also induced during the isolation procedure. The presenting molecule for the stress antigen is unknown. It appears to be a molecule with little or no polymorphism (Havran *et al.*, 1991; Allison and Havran,

1991).

A similar response is generated by i-IEL in vitro. The T cells and epithelial cells of the small intestine can be isolated from the gut and separated. Adding these 2 cell populations back together results in activation of the i-IEL. Soluble anti$\gamma\delta$ antibodies block the response, indicating that the response is specific. AntiTL antibodies also inhibit the response, suggesting that the response is TL restricted (see above). Other cells can not substitute for the epithelial cells, showing that the interaction is specific. The $\gamma\delta$ cells responsible for the activity are probably the V$\gamma5^+$ i-IEL, which constitute 70-90% of all $\gamma\delta$ cells in the gut epithelium. Heat shock increases the response. Thus, the i-IEL are autoreactive cells, which probably respond to a stress protein of the small intestine epithelial cells which is presented by TL antigen(s) (Eghtesady and Kronenberg, 1992; Eghtesady *et al.*, 1992).

A major difference between the DEC and i-IEL populations is that the DEC $\gamma\delta$ TCR is identical in all cells, whereas the i-IEL Vγ and Vδ genes and their junctions are not (see above). The Vγ5 gene is used by most i-IEL $\gamma\delta$ cells, but the VJ junction can vary. In addition, there are at least 5 Vδ genes present in the i-IELs (see above, table II), which makes for extensive overall diversity.

Why do i-IEL have this diversity, whereas the DECs do not?

Several possible explanations come to mind. A simple mechanistic reason may acount (at least in part) for the difference: the DEC precursor cell develop in the fetal thymus at a time when little terminal transferase activity is present, which effectively limits the amount of junctional diversity.

An alternative reason may be that the gut epithelium cells may present several different self peptides to different populations of γδ cells. Or the gut epithelium can present the self peptide(s) on several different TL molecules, and each combination is recognized by i-IELs with different γδ TCRs. Alternatively, the V(D)J junctions may be of low importance in the recognition of the ligand. The V region outside the junction may primarily determine the binding affinity. Or the difference may be related to the ways in which these 2 γδ subsets develop. The DECs must traverse through the thymus where the cells with the canonical junctions are (probably) positively selected, allowed to exit and populate the skin (Itohara and Tonegawa, 1990b). The Vγ5 i-IEL are present in nude mice and in ß2m deficient mice, which indicates that they do not need the thymus per se and that they are not selected on class I/Ib molecules. Hence without (positive) selection, the emerging cells are expected to express variable junctions. However, the prediction then is that only those with the 'correct'

junctions will respond to stressed epithelial cells (if the junctions influence the specificity). If in vivo, as in vitro, the cells do not proliferate well, clonal expansion does not occur and a Vγ5 subset with homogeneous joints will not be observed.

It may be possible to investigate this latter hypothesis. One can generate $V\gamma5^+$ hybridomas and those that respond to stressed gut epithelium can be analyzed for V gene usage and junctional diversity.

In summary, the Vγ3/Vδ1 subset develops early in the fetal thymus and localizes to the skin. The V(D)J junctions of virtually all cells are identical. These DECs respond to stressed keratinocytes. The Vγ4/Vδ1 subset develops after the Vγ3/Vδ1 subset, also has invariant TCRs and is found in the female reproductive organs as well. The specificity of these cells is unknown. The Vγ2, Vγ1.1 and Vγ1.2 subsets have heterogeneous junctions and can be found in the spleen and lymph nodes. The Vγ1.1 subset appears to be involved in responses to mycobacterial antigens and possibly hsp.

The Vγ5 subset makes up the majority of the γδ cells in the small intestine epithelium. The junctions of these cells are diverse and the cells do not need to traffic through the thymus. These cells respond to stressed epithelial cells of

the small intestine, a reaction which appears to be restricted by TL molecules. Very few cell surface molecules that can serve as ligands for γδ cells have been defined and those that are known are either class I, Ib, II or class I like (and TCT.1). Generating γδ cells that are specific for these structures appears to be difficult. Few specific antigens or peptides which can stimulate specific γδ cells are known. Superantigens which can stimulate large subsets of γδ cells appear to exist, at least in humans.

STRUCTURE OF THE γ CHAIN.

Protein structure.

The murine γδ TCR is expressed as a disulphide linked heterodimer expressed on the surface of T cells. The size of this complex varies between 77 and 90 kD, depending on the particular γ and δ chains that are used. On a denaturing gel, the γ chain migrates as a protein of between 31 and 47 kD, depending on the particular γ constant region which is used. The size of the δ chain varies between 45 and 48 kD. BALB/c mice have genes 4 Cγ regions. These are designated Cγ1 through Cγ4, with Cγ3 being a pseudogene (see below).

Immunoprecipitation of TCR γ chains from C57BL/10 and BALB/c spleen cells identified 3 types of γ chains as shown in Table VI.

Table IV. Sizes of the TCR γ chains, as determined by immunoprecipitation.

	size (kD)	
γ chain	BALB/c	C57BL/10
V2J1Cγ1	35	32 (no CHO)
V1.2J2Cγ2	31 (no CHO)	31 (no CHO)
V1.1J4Cγ4	41	41

No CHO: no carbohydrates; from Cron *et al.*, 1990.

The Cγ2 chains are not glycosylated. This allows for easy identification of Cγ1, 2 and 4 containing γ chains in mouse strains like BALB/c, where Cγ1 is N-linked glycosylated (see table VI). TCR γ chains containing Cγ4 are larger than those with Cγ1 or Cγ2, due to the larger protein core. This γ chain is glycosylated in all mouse strains examined. In most strains the V1.1J4Cγ4 protein is 41 kD (table VI), of which 6 kD is due to carbohydrates (Cron *et al.*, 1990). In a few mouse strains the Cγ4 containing γ chain is larger (47 kD), due to a larger protein core size (Cron *et al.*, 1990).

Genomic organization of the murine γ locus.

The TCR γ chains are encoded by gene segments which are located in the TCR γ locus. BALB/c mice have 7 γ V gene segments, 4 J gene segments and 4 constant region genes. The

organization of the murine γ locus is unusual, in that each of the 4 C regions, has its own V gene(s). This is different from most other mouse TCR and Ig loci. Four gene clusters have been identified, each with a single C and J gene segment and a variable number (0-4) of V gene segments. Figure 4 shows a schematic view of how these gene clusters are believed to be organized in BALB/c mice. The clusters were first proposed when it became apparent that there was a preferential rearrangement pattern. The V genes in each cluster rearrange most frequently or exclusively to the C in the same cluster (Raulet *et al.*, 1989).

The Cγ1 cluster contains Vγ2, 3, 4 and 5. Rearrangements within this cluster to Jγ1 have established the order of the V genes as shown in figure 1 (Vγ5-2-4-3--J-C). The Cγ3 cluster contains only Vγ1.3, whereas cluster Cγ2 contains no V genes. The Cγ4 cluster contains Vγ1.1 and 1.2. Since Vγ1.2 is only found rearranged to Jγ2-Cγ2 and since this V gene is in the opposite orientation (rel. to Vγ1.1 and Cγ4), it is likely that the Cγ2 cluster is next to the Cγ4 cluster, oriented as indicated in figure 4. In rare Vγ5 to Cγ4 rearrangements, all other V and C genes are deleted (Pelkonen *et al.*, 1987). The results of these rearrangement established the order of the clusters as indicated in figure 4. Gaps in the map are located between the clusters between Vγ2 and 4, and between Vγ3 and Jγ1 (see figure 4; Hayday *et*

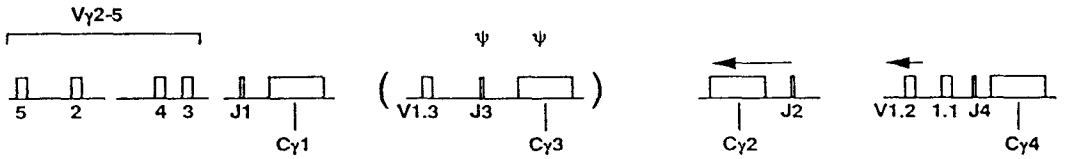


Figure 4. Proposed genomic organization of the murine TCR γ locus. The orientation of the Cγ3 cluster is unknown, as the parentheses indicate. Arrows show the transcriptional orientation of the Vγ1.2, Jγ2 and Cγ2 gene segments (see text) and ψ indicate pseudo gene segments. The map is not to scale (adapted from Raulet, 1989).

al., 1985; Garman *et al.*, 1986; Traunecker *et al.*, 1986; Pelkonen *et al.*, 1987; Iwamoto *et al.*, 1986).
Field inversion gel electrophoresis (FIGE) has confirmed that the Cγ2 and Cγ4 clusters are linked on a single Sal I fragment (90 kb) and that the Cγ3 cluster is within 60 kb of this fragment. The Cγ1 cluster was not linked to the other clusters (Woolf *et al.*, 1988).

Most other TCR and Ig loci have an organization with one or more C regions, all at one end of the locus, a number of J segments upstream of the C region(s), D regions upstream of the Js (only in Ig heavy chain, TCR ß and δ loci) and multiple Vs at the other end of the locus, upstream of the

Ds/Js (Lai *et al.*, 1989).

The mouse γ locus is also different from the human TCR γ locus in organization. The organization of the human locus is similar to that of the other TCR loci (Lefranc and Rabbitts, 1989). It is schematically shown in figure 5.

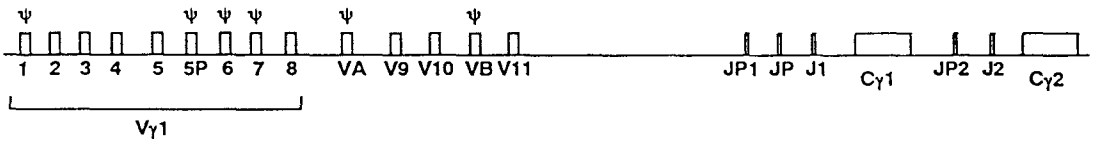


Figure 5. Genomic organization of the human TCR γ locus. Pseudogenes are indicated by ψ. Individual exons are not shown (adapted from Lefranc and Rabbitts, 1989).

The mouse γ locus is similar only to the mouse Ig lambda light chain locus and the shark Ig heavy chain (IgH) loci. The mouse lambda light chain locus has 2 C regions, each with 2 C genes and 1 or 2 V genes (Lai *et al.*, 1989). The shark IgH locus also resembles this organization. It has gene clusters containing one V, one J, one or more Ds and one C region. These clusters are repeated many times in the shark genome (Kokubu *et al.*, 1988).

Most of the productive rearrangements which give rise to cell surface expression of γ chains involve intra cluster

rearrangement, i.e., Vγ5-4-3-2 rearrange to Jγ1, Vγ1.1 rearranges only to Jγ4 and Vγ1.2 only to Jγ2. However, rearrangements between clusters are occasionally observed, but they are usually nonfunctional (Raulet, 1989).

γ gene sequences.

C genes. The constant region genes encode the constant region of the γ protein chain. The Cγ1, 2 and 3 genes in the mouse consist of 3 exons, which encode the extracellular domain (330 nucleotides; nt.), most of the hinge region (30-45 nt.), the transmembrane region and the cytoplasmic domain (140 nt) plus untranslated region. The 3' untranslated regions (from the stop codon to the polyadenylation signal) are 398/406 and >283 nt. for Cγ1/Cγ2/Cγ3 and Cγ4 (Hayday *et al.*, 1985; Garman *et al.*, 1986; Iwamoto *et al.*, 1986). Sequence comparison shows that Cγ1, Cγ2 and Cγ3 are very homologous (over 90% at the nucleotide level, Garman *et al.*, 1986, chapter III). However, Cγ3 appears to be nonfunctional, due to a defective splice site. It also has a mutation in the polyadenylation signal (Hayday *et al.*, 1985). The upstream J segment (Jγ3) is also nonfunctional due to a stop codon in the coding sequence (Traunecker *et al.*, 1986). That these gene segments are nonessential is clear since both (and Vγ1.3) are absent in several mouse strains (Iwamoto *et al.*, 1986; Klotz *et al.*, 1989).

The Cγ1 gene in BALB/c encodes a site for N glycosylation in its first exon, which is not present in Cγ2 (see table VI; Garman *et al.*, 1986).

Cγ4 is also homologous to Cγ1, 2 and 3, but less than the latter 3 Cγ genes among one another (only 66% at the nucleotide level). The Cγ4 gene is most divergent from the other mouse Cγ genes in the hinge/transmembrane exon, in length and sequence composition. The homology in the other 2 exons is substantially higher (77%).

J gene segments. The 4 J gene segments are homologous, with Jγ4 being the most diverse. The Jγ gene segments encode 19 amino acids. Jγ3 is a pseudogene, as it contains a stop codon. The recognition sequences 5' to the Jγs contain a 12 bp spacer (Hayday *et al.*, 1985; Traunecker *et al.*, 1986).

V gene segments. Seven Vγ gene segments have been described in Balb/c mice. The Vγ1.1, 1.2 and 1.3 genes are very similar (>94%) and make up the Vγ1 family. Vγ2 is homologous to the Vγ1 family members (65-70% at the nucleotide level). The other 3 Vγs show limited similarity to one another and to the Vγ1 gene segments (Huck *et al.*, 1988; chapter III). The Vγ2, 3, 4 and 5 gene segments are all located 5' to Cγ1, whereas Vγ1.3 is the single V 5' to Jγ3-Cγ3. Vγ1.1 and 1.2 are located between Cγ4 and Cγ2 (see figure 4). The

recognition sequences 3' to the V genes contain 23 bp spacers (Hayday *et al.*, 1985; Garman *et al.*, 1986).

In order to better understand the organization of the murine TCR γ locus and to understand how this organization may have evolved, the locus was characterized in detail and compared to the structure of the human TCR γ locus. Chapter 2 describes the organization of the murine γ locus and the position of the γ gene elements, including 2 new elements which are potential γ specific enhancers.

The γ gene sequences of various vertebrates were compared as well in order to determine which γ gene segments were conserved during evolution and therefore may have an essential function. The sequence comparisons also indicate how the γ loci may have evolved differently in the mouse and human lineages. Chapter 3 shows these sequence comparisons, describes the conserved structural elements and describes the evolutionary implications.

REFERENCES.

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

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

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

Augustin, A., Kubo, R. T., and Sim, G. (1989). Resident pulmonary lymphocytes expressing the gamma/delta T-cell receptor. Nature *340*, 239-241.

Band, H., Porcelli, S. A., Panchamoorthy, G., McLean, J., Morita, C. T., Ishikawa, S., Modlin, R. L., and Brenner, M. B. (1991). Antigens and antigen-presenting molecules for gamma/delta T cells. Curr. T. Micr. *173*, 229-234.

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 gamma-delta. Proc. Natl. Acad. Sci. USA *88*, 43-47.

Barrett, T. A., Delvy, M. L., Kennedy, D. M., Lefrancois, L., Matis, L. A., Dent, A. L., Hedrick, S. M., and Bluestone, J. A. (1992). Mechanism of self-tolerance of gamma/delta T cells in epithelial tissue. J. Exp. Med. *175*, 65-70.

Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L., and Wiley, D. C. (1987). The foreign antigen binding site and T cell recognition regions ao class I histocompatibility antigens. Nature *329*, 512-518.

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

Bluestone, J. A., Cron, R. Q., Cotterman, M., Houlden, B. A., and Matis, L. A. (1988). Structure and specificity of T cell receptor gamma/delta on major histocompatibity complex antigen-specific

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 gamma-delta cells. Immunol. Rev. *120*, 5-33.
CD3+, CD4-, CD8- T lymphocytes. J. Exp. Med. *168*, 1899-1916.

Bonneville, M., Janeway Jr, C. A., Ito, K., Haser, W., Ishida, I., Nakanishi, N., and Tonegawa, S. (1988). Intestinal intraepithelial lymphcytes are a distinct set of gamma-delta T cells. Nature *336*, 479-481.

Bonneville, M., Ito, K., Krecko, E. G., Itohara, S., Kappes, D., Ishida, I., Kanagawa, O., Janeway Jr, C. A., and Murphy, D. B. (1989). Recognition of a self major histocompatibility complex TL region product by gamma/delta T-cell receptors. Proc. Natl. Acad. Sci. USA *86*, 5928-5932.

Bonneville, M., Itohara, S., Krecko, E. G., Mombaerts, P., Ishida, I., Katsuki, M., Berns, A., Farr, A., Janeway, C. A., and Tonegawa, T. (1990). Transgenic mice demonstrate that epithelial homing of gamma/delta T cells is determined by cell lineages independent of T cell receptor specificity. J. Exp. Med. *171*, 1015-1026.

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 gamma/delta T lymphocytes. Science *249*, 67-69.

Bosnes, V., Qvigstad, E., Lundin, K. E. A., and Thorshy, E. (1990). Recognition of a particular HLA-DQ heterodimer by a human gamma/delta T cell clone. Eur. J. Immunol. *20*, 1429-1433.

Bucy, R. P., Chen, C.-L., Cihak, J., Loesch, U., and Cooper, M. D. (1988). Avian T cells expressing gamma/delta receptors localize in the splenic sinusoids and the intestinal epithelium. J. Immunol. *141*, 2200-2205.

Carding, S. R., Kyes, S., Jenkinson, E. J., Kingston, R., Bottomly, K., Owen, J. J. T., and Hayday, A. C. (1990). Developmentally regulated fetel thymic and extrathymic T-cell receptor gamma-delta gene expression. Genes and Dev. *4*, 1304-1315.

Cheng, S. H., Penninger, J. M., Ferrick, D. A., Molina, T. J., Wallace, V. A., and Mak, T. W. (1991). Biology of murine gamma-delta T cells. Crit. Rev. Immunol. *11*, 145-166.

Ciccone, E., Viale, O., Pende, D., Malnati, M., Ferrara, G. B., Moretta, A., and Moretta, L. (1989). Specificity of human T lymphocytes expressing a gamma/delta T cell antigen receptor. Recognition of a polymorphic determinant of HLA class I molecules by a gamma/delta clone. Eur. J. Immunol. *19*, 1267-1271.

Cooper, M. D., Bucy, R. P., Lahti, J. M., Char, D., and Chen, C.-L. H. (1989). T-cell development in birds. Cold Spring Harbor Symp. Quant. Biol *LIV*, 69-73.

Cron, R. Q., Coligan, J. E., and Bluestone, J. A. (1990). Polymorphisms and diversity of T-cell receptor-gamma proteins expressed in mouse spleen. Immunogenetics *31*, 220-228.

Cron, R. Q., Koning, F., Maloy, W. L., Pardoll, D., Coligan, J. E., and Bluestone, J. A. (1988). Peripheral murine CD3+, CD4- CD8- T lymphocytes express novel T cell receptor gamma delta structures. J. Immunol. *141*, 1074-1082.

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

Del Porto, P., Mami-Chouaib, F., Bruneau, J., Jitsukawa, S.,

Dumas, J., Harnois, M., and Hercend, T. (1991). TCT.1, a target molecule for gamma/delta T cells, is encoded by an immunoglobulin superfamily gene (Blast-1) located in the CD1 region of human chromosome 1. J. Exp. Med. *173*, 1339-1344.

Deusch, K., Luling, F., Reich, K., Classen, M., Wagner, H., and Pfeffer, K. (1991). A major fraction of human intraepithelial lymphocytes simultaneously expresses the gamma/delta T cell receptor, the CD8 accessory molecule and preferentially uses the V-delta-1 gene segment. Eur. J. Immunol. *21*, 1053-1059.

Eghtesady, P., and Kronenberg, M. (1992). Intestinal γδ T lymphocytes are autoreactive for stressed intestinal epithelial cells. Submitted.

Eghtesady, P., Panwala, C., Teitell, M. and Kronenberg, M. (1992). Recognition of the thymus leukemia (TL) antigen by intestinal epithelial γδ T lymphocytes. Submitted.

Elbe, A., Tschachler, E., Steiner, G., Binder, A., Wolff, K., and Stingl, G. (1989). Maturational steps of bone marrow-derived dendritic murine epidermal cells. J. Immunol. *143*, 2431-2438.

Elliott, J. F., Rock, E. P., Patten, P. A., Davis, M. M., and Chien, Y. (1988). The adult T-cell receptor delta-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., and Coligan, J. E. (1990). T cell receptor delta-gene expression and diversity in the mouse spleen. J. Immunol. *145*, 1311-1317.

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

Fisch, P., Malkovsky, M., Kovats, S., Sturm, E., Braakman, E., Klein, B. S., Voss, S. D., Morrissey, L. W., DeMars, R., Welch, W. J., Bolhuis, R. L. H., and Sondel, P. M. (1990). Recognition by human V-gamma-9/V-delta-2 T cells of a GroEL homolog on Daudi Burkitt's lymphoma cells. Science *250*, 1271-1273.

Fisch, P., Kovats, S., Fundim, N., Strum, E., Braakman, E., DeMars, R., Bolhuis, R. L. H., Sondel, P. M., and Malkovsky, M. (1991). Function and specificity of human

V-gamma-9/V-delta-2 T lymphocytes. Curr. T. Micr. *173*, 179-182.

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

Griffin, J. P., Harshan, K. V., Born, W. K., and Orme, I. M. (1991). Kinetics of accumulation of gamma/delta receptor-bearing T lymphocytes in mice infected with live mycobacteria. Infect. Imm. *59*, 4263-4265.

Groh, V., Porcelli, S., Fabbi, M., Lanier, L. L., Picker, L. J., Anderson, T., Warnke, R. A., Bhan, A. K., Strominger, J. L., and Brenner, M. B. (1989). Human lymphocytes bearing T cell receptor gamma/delta are phenotypically diverse and evenly distributed throughout the lymphoid system. J. Exp. Med. *169*, 1277-1294.

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. J. Exp. Med. *173*, 471-481.

Guy-Grand, S., and Vassalli, P. (1986). Gut injury in mouse graft-versus-host reaction study of its occurrence and mechanisms. J. Clin. Invest. 77, 1584-1595.

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 gamma genes during thymic development. J. Exp. Med. *164*, 1-24.

Happ, M. P., Kubo, R. T., Palmer, E., Born, W. K., and O'Brien, R. L. (1989). Limited receptor repertoire in a mycobacteria-reactive subset of gamma/delta T lymphocytes. Nature *342*, 696-698.

Haregewoin, A., Soman, G., Hom, R. C., and Finberg, R. W. (1989). Human gamma/delta T cells respond to mycobacterial heat-shock protein. Nature *340*, 309-312.

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

Havran, W. L., and Allison, J. P. (1990). Origin of Thy-1+ dendritic epidermal cells of adult mice from fetal thymic precursors. Nature *344*, 68-70.

Havran, W. L., Chien, Y., and Allison, J. P. (1991). Recognition of self antigens by skin-derived T cells with invariant gamma/delta antigen receptor. Science *252*, 1430-1432.

Havran, W. L., Poenie, M., Tigelaar, R. E., Tsien, R. Y., and Allison, J. P. (1989). Phenotypic and functional analysis of gamma-delta T cell receptor-positive murine dendritic epidermal clones. J. Immunol. *142*, 1422-1428.

Hayday, A. C., Saito, H., Gillies, D., Kranz, D. M., Tanigawa, G., Eisen, H. N., and Tonegawa, S. (1985). Structure, Organization, and somatic rearrangement of T cell gamma genes. Cell *40*, 259-269.

Hein, W. R., and Mackay, C. R. (1991). Prominence of gamma-delta T cells in the ruminant immune system. Imm. Today 12, 30-34.

Hershberg, R., Eghtesady, P., Sydora, B., Brorson, K., Cheroutre, H., Modlin, R., and Kronenberg, M. (1990). Expression of the thymus leukemia antigen in mouse intestinal epithelium. Proc. Natl. Acad. Sci. USA *87*, 9727-9731.

Holoshitz, J., Koning, F., Coligan, J. E., De Bruyn, J., and Strober, S. (1989). Isolation of CD4- CD8- mycobacteria-reactive T lymhocyte clones from rheumatic arthritis synovial fluid. Nature *339*, 226-229.

Houlden, B. A., Cron, R. Q., Coligan, J. E., and Bluestone, J. A. (1988). Systematic development of distinct T cell receptor-gamma delta T cell subsets during fetal ontogeny. J. Immunol. *141*, 3753-3759.

Huck, S., Dariavach, P., and LeFranc, M. P. (1988). Variable region genes in the human T-cell rearranging gamma (TRG) locus: V-J junction and homology with the mouse genes. Embo. J. 7, 719-726.

Ikuta, K., Kina, T., MacNeil, I., Uchida, N., Peault, B., Chien, Y., and Weissman, I. L. (1990). A developmental switch in thymic lymphocyte maturation potential occurs at the level of hematopoietic stem cells. Cell *62*, 863-874.

Ikuta, K., and Weissman, I. L. (1991). The junctional modifications of a T-cell receptor gamma chain are determined at the level of thymic precursors. J. Exp. Med. *174*, 1279-1282.

Ito, K., Bonneville, M., Takagaki, Y., Nakanishi, N., Kanagawa, O., Krecko, E. G., and Tonegawa, S. (1989). Different gamma delta T-cell receptors are expressed on thymocytes at different stages of development. Proc. Natl. Acad. Sci. USA *86*, 631-635.

Ito, K., Van Kaer, L., Bonneville, M., Hsu, S., Murphy, D. B., and Tonegawa, S. (1990). Recognition of the product of a novel MCH TL region gene (27b) by a mouse gamma/delta T cell receptor. Cell *62*, 549-561.

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. Proc. Natl. Acad. Sci. USA *86*, 5094-5098.

Itohara, S., Farr, A. G., Lafaille, J. J., Bonneville, M., Takagaki, Y., Werner, H., and Tonegawa, S. (1990a). Homing of a gamma-delta thymocyte subset with homegeneous T-cell receptors to mucosal epithelia. Nature *343*, 754-757.

Itohara, S., and Tonegawa, S. (1990b). Selection of gamma-delta cells with canonical T-cell antigen receptors in fetal thymus. Proc. Natl. Acad. Sci. USA *87*, 7935-7938.

Iwamoto, A., Rupp, F., Ohashi, P. S., Walker, C. L., Pircher, H., Joho, R., Hengartner, H., and Mak, T. W. (1986). T cell-specific gamma genes in C57BL/10 mice sequence and expression of new constant and variable region genes. J. Exp. Med. *163*, 1203-1212.

Janeway, C. A. (1989). A primitive immune system. Nature 341, 108.

Janis, E. M., Kaufmann, S. H. E., Schwartz, R. H., and Pardoll, D. M. (1989). Activation of gamma/delta T cells in the primary immune response to mycobacterium tuberculosis. Science *244*, 713-716.

Jarry, A., Cerf-Bensussan, N., Brousse, N., Selz, F., and Guy-Grand, D. (1990). Subsets of CD3+ (T cell receptor alpha/beta or gamma/delta) and CD3- lymphocytes isolated from normal human gut epithelium display phenotypical features different from their counterparts in peripheral blood. Eur. J. Immunol. *20*, 1097-1103.

Jitsukawa, S., Triebel, F., Faure, F., Miossec, C., and Hercend, T. (1988). Cloned CD3+ TcR alpha/beta- Ti-gamma-A- peripheral blood lymphocytes compared to the Ti-gamma-A+ counterparts: structural differences of the gamma/delta

receptor and functional heterogeneity. Eur. J. Immunol. *18*, 1671-1679.

Johnson, R. M., Lancki, D. W., Sperling, A. I., Dick, R. F., Spear, P. G., Fitch, F. W., and Bluestone, J. A. (1992). A murine CD4- CD8- Tcell receptor-gamma/delta T lymphocyte clone specific for herpes simplex virus glycoprotein. J. Immunol. *148*, 983-988.

Kabelitz, D., Bender, A., Schondelmaier, S., Schoel, B., and Kaufmann, S. H. E. (1990). A large fraction of human peripheral blood $\gamma\delta^+$ T cells is activated by *Mycobacterium tuberculosis* but not by its 65-kD heat shock protein. J. Exp. Med. *171*, 667-679.

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. J. 7, 3413-3422.

Koning, F., Yokoyama, W. M., Maloy, W. L., Stingl, G., McConnell, T. J., Cohen, D. I., Shevach, E. M., and Coligan, J. E. (1988). Expression of C-gamma-4 T cell receptors and lack of isotype exclusion by dendritic epidermal T cell lines. J. Immunol. *141*, 2057-2062.

Korman, A. J., Marusic-Galesic, S., Spencer, D., Kruisbeek, A. M., and Raulet, D. H. (1988). Predominant variable region gene usage by gamma/delta T cell receptor-bearing cells in the adult thymus. J. Exp. Med. *168*, 1021-1040.

Kozbor, D., Trinchieri, G., Monos, D. S., Isobe, M., Russo, G., Haney, J. A., Zmijewski, C., and Croce, C. M. (1989). Human TCR-gamma+/delta+, CD8+ T lymphocytes recognize tetanus toxoid in an MHC-restricted fashion. J. Exp. Med. *169*, 1847-1851.

Kyes, S., Carew, E., Carding, S. R., Janeway Jr., C. A., and Hayday, A. (1989). Diversity in T-cell receptor gamma gene in intestinal epithelium. Proc. Natl. Acad. Sci. USA *86*, 5527-5531.

Kyes, S., and Hayday, A. (1990). Disparate types of gamma-delta T cell. Res. in Immunol. *141*, 583-587.

Lacy, M. J., McNeil, L. K., Roth, M. E., and Kranz, D. M. (1988). T-cell receptor delta-chain diversity in peripheral lymphocytes. Proc. Natl. Acad. Sci. USA *86*, 1023-1026.

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

Lai, E., Wilson, R. K., and Hood, L. (1989). Physical maps of the mouse and human immunoglobulin-like loci. Adv. Immunol. *46*, 1-59.

Lefranc, M., and Rabbitts, T. H. (1989). The human T-cell receptor gamma (TRG) genes. TIBS *14*, 214-218.

Lefrancois, L., LeCorre, R., Bluestone, J. A., and Goodman, T. (1991). Selection of V-delta+ T cell receptors of intestinal intraepithelial lymphocytes is dependent on class II histocompatibility antigen expression. Curr. T. Micr. *173*, 255-267.

Lefrancois, L., LeCorre, R., Mayo, J., Bluestone, J. A., and Goodman, T. (1990). Extrathymic selection of TCR

gamma-delta+ T cells by class II major histocompatibiliy complex molecules. Cell *63*, 333-340.

Maeda, K., Nakanish, N., Rogers, B. L., Haser, W. G., Shitara, K., Yoshida, H., Takagaki, Y., Augustin, A. A., and Tonegawa, S. (1987). Expression of the T-cell receptor gamma-chain gene products on the surface of peripheral T cells and T-cell blasts generated by allogeneic mixed lymphocyte reaction. Proc. Natl. Acad. Sci. USA *84*, 6536-6540.

Mami-Chouaib, F., Del Porto, P., Delorme, D., and Hercend, T. (1991). Further evidence for a gamma/delta T cell receptor-mediated TCT.1/CD48 recognition. J. Immunol. *147*, 2864-2867.

Matis, L. A., Cron, R., and Bluestone, J. A. (1987). Major histocompatibility complex-linked specificity of gamma/delta receptor-bearing T lymphocytes. Nature *330*, 262-264.

Matis, L. A., Fry, A. M., Cron, R. Q., Cotterman, M. M., Dick, R. F., and Bluestone, J. A. (1989). Structure and specificity of Class II MHC alloreactive gamma/delta T cell receptor heterodimer. Science *245*, 746-749.

McConnel, T. J., Yokoyama, W. M., Kikuchi, G. E., Einhorn, G. P., Stingl, G., Shevach, E. M., and Coligan, J. E. (1989). Delta-chains of dendritic epidermal T cell receptors are diverse but pair with gamma-chains in a restricted manner. J. Immunol. *142*, 2924-2931.

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. J. Immunol. *145*, 1369-1375.

Nandi, D., and Allison, J. P. (1991). Phenotypic analysis and gamma-delta T cell receptor repertoire of murine T cells associated with the vaginal epithelium. J. Immunol. *147*, 1773-1778.

O'Brien, R. L., Happ, M. P., Dalles, A., Palmer, E., and Kubo, R. (1989). Stimulation of a major subset of lymphocytes expressing T cell receptor gamma/delta by an antigen derived from mycobacterium tuberculosis. Cell *57*, 667-674.

O'Brien, R. L., and Born, W. (1991). Specificity of mycobacteria/self-reactive gamma/delta cells. Curr. T. Micr.

173, 143-150.

O'Brien, R. L., Happ, M. P., Dalles, A., Cranfill, R., Hall, L., Lang, J., Fu, Y., Kubo, R., and Born, W. (1991). Recognition of a single hsp-60 epitope by an entire subset of gamma/delta T lymphocytes. Immunol. Rev. *121*, 156-170.

Ogimoto, M., Yoshikai, Y., Matsuzaki, G., Matsuzaki, K., Kishihara, K., and Nomoto, K. (1990). Expression of T cell receptor V-gamma-5 in the adult thymus of irradiated mice after transplantation with fetal liver cells. Eur. J. Immunol. *20*, 1965-1970.

Ohmen, J. D., Barnes, P. F., Uyemura, K., Lu, S., Grisso, C. L., and Modlin, R. L. (1991). The T cell receptors of human gamma/delta T cells reactive to mycobacterium tuberculosis are encoded by specific V genes but diverse V-J junctions. J. Immunol. *147*, 3353-3359.

Pardoll, D. M., Fowlkes, B. J., Lew, A. M., Maloy, W. L., Weston, M. A., Bluestone, J. A., Schwartz, R. H., Coligan, J. E., and Kruisbeek, A. M. (1988). Thymus-dependent and thymus-independent developmental pathways for peripheral T cell receptor-gamma-delta-bearing lymphocytes. J. Immunol. *140*, 4091-4096.

Parker, C. M., Groh, V., Band, H., Porcelli, S. A., Morita, C., Fabbi, M., Glass, D., Strominger, J. L., and Brenner, M. B. (1990). Evidence for extrathymic changes in the T cell receptor gamma/delta repertoire. J. Exp. Med. *171*, 1579-1612.

Payer, E., Elbe, A., and Stingl, G. (1991). Circulating CD3+/T cell receptor V-gamma-3+ fetal murine thymocytes home to the skin and give rise to proliferating dendritic epidermal T cells. J. Immunol. *146*, 2536-2543.

Pelkonen, J., Traunecker, A., and Karjalainen, K. (1987). A new mouse TCR V-gamma gene that shows remarkable evolutionary conservation. Embo. J. *6*, 1941-1944.

Pfeffer, K., Schoel, B., Plesnila, N., Lipford, G. B., Kromer, S., Deusch, K., and Wagner, H. (1992). A lectin-binding, protease-resistant mycobacterial ligand specifically activates V-gamma-9+ human gamma/delta T cells. J. Immunol. *148*, 575-583.

Porcelli, S., Brenner, M. B., Greenstein, J. L., Balk, S. P., and Terhorst, C. (1989). Recognition of cluster of differentiation 1 antigens by human CD48-CD8- cytolytic T lymphocytes. Nature *341*, 447-450.

Rajasekar, R., Sim, G., and Augustin, A. (1990). Self heat shock and gamma/delta T-cell reactivity. Proc. Natl. Acad. Sci. USA *87*, 1767-1771.

Raulet, D. H. (1989). The structure, function, and molecular genetics of the gamma/delta T cell receptor. Ann. Rev. Immunol. *7*, 175-207.

Rellahan, B. L., Bluestone, J. A., Houlden, B. A., Cotterman, M. M., and Matis, L. A. (1991). Junctional sequences influence the specificity of gamma/delta T cell receptors. J. Exp. Med. *173*, 503-506.

Rust, C. J. J., Verreck, F., Vietor, H., and Koning, F. (1990). Specific recognition of staphylococcal enterotoxin A by human T cells bearing receptors with the V-gamma-9 region. Nature *346*, 572-574.

Salter, R. D., Benjamin, R. J., Wesley, P. K., Buxton, S. E., Garrett, P. T. J., Clayberger, C., Krensky, A. M., Norment, A. M., Littman, D. R., and Parham, P. (1990). A binding site for the T-cell coreceptor CD8 on the alpha3 domain of HLA-A2. Nature, *345*, 41-46.

Sim, G., and Augustin, A. (1990). Dominantly inherited

expression of BID, an invariant undiversified T cell receptor delta chain. Cell *61*, 397-405.

Sim, G., and Augustin, A. (1991). Extrathymic positive selection of gamma delta T cells V-gamma-4 J-gamma-1 rearrangements with "GxYS" junctions. J. Immunol. *146*, 2439-2445.

Spits, H., Paliard, X., Engelhard, V. H., and De Vries, J. E. (1990). Cytotoxic activity and lymphokine productio of T cell receptor (TCR)-alpha/beta+ and TCR-gamma/delta+ cytotoxic T lymphocyte (CTL) clones recognizing HLA-A2 and HLA-A2 mutants. Recognition of TCR-gamma/delta+ CTL clones is affected by mutations at positions 152 and 156. J. Immunol. *144*, 4156-4162.

Staunton, D. E., and Thorley-Lawson, D. A. (1987). Molecular cloning of the lymphocyte activation marker Blast-1. Embo. J. *6*, 3695-3701.

Strominger, J. L. (1989). The gamma/delta T cell receptor and class lb mhc-related proteins: enigmatic molecules of immune recognition. Cell *57*, 895-898.

Stroynowski, I. (1990). Molecules related to class-I major

histocompatibility complex antigens. Ann. Rev. Immunol. *8*, 501-530.

Sturm, E., Braakman, E., Fisch, P., Sondel, P. M., Bolhuis, R. L. H. (1991). Daudi cell specificity correlates with the use of V-gamma-9 V-delta-2 encoded TCR-gamma-delta. Curr. T. Micr. 173, 183-188.

Takagaki, Y., DeCloux, A., Bonneville, M., and Tonegawa, S. (1989a). Diversity of gamma-delta T-cell receptors on murine intestinal intraepithelial lymphocytes. Nature *339*, 712-714.

Takagaki, Y., Nakanishi, N., Ishida, I., Kanagawa, O., and Tonegawa, S. (1989b). T cell receptor-gamma genes preferentially utilized by adult thymocytes for the surface expression. J. Immunol. *142*, 2112-2121.

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

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 gamma chain. Embo. J. *5*, 1589-1593.

Ullrich, R., Schieferdecker, H. L., Ziegler, K., Riecken, E. O., and Zeitz, M. (1990). Short communications gamma/delta cells in the human intestine express surface markers of activation and are preferentially located in the epithelium. Cell. Immunol. *128*, 619-627.

Vidovic, D., Roglic, M., McKune, K., Guerder, S., MacKay, C., and Dembic, Z. (1989). Qa-1 restricted recognition of foreign antigen by a gamma/delta T-cell hybridoma. Nature *340*, 646-650.

Vodovic, D., and Dembic, Z. (1991). Qa-1 restricted gamma/delta T cells can help B cells. Curr. T. Micr. *173*, 239-244.

Whetsell, M., Mosley, R. L., Whetsell, L., Schaefer, F. V., Miller, K. S., and Klein, J. R. (1991). Rearrangement and junctional-site sequence analyses of T-cell receptor gamma genes in intestinal intraepithelial lymphocytes from murine athymic chimeras. Mol. Cell. Biol. *11*, 5902-5909.

Woolf, T., Lai, E., Kronenberg, M., and Hood, L. (1988).

Mapping genomic organization by field inversion and two-dimensional gel electrophoresis: Application to the murine T-cell receptor gamma gene family. Nucleic Acids Res. *16*, 3863-3875.

Wu, M., Van Kaer, L., Itohara, S., and Tonegawa, S. (1991). Highly restricted expression of the thymus leukemia antigen on intestinal epithelial cells. J. Exp. Med. *174*, 213-218.

Yoshikai, Y., Reis, M. D., and Mak, T. W. (1986). Athymic mice express a high level of functional gamma-chain but greatly reduced levels of alpha- and beta-chain T-cell receptor messages. Nature *324*, 482-485.

Yoshikai, G., Matsuzaki, G., Inoue, T., and Nomoto, K. (1990). An increase in number of T-cell receptor gamma/delta bearing T cells in athymic nude mice treated with complete Freund's adjuvants. Immunol. Today *70*, 61-65.

Chapter II.

CLONING OF THE MURINE T-CELL RECEPTOR GAMMA LOCUS.

Complete genomic organization, 2 new potential enhancers and a Cγ4 gene segments with multiple hinge encoding exons.

Bernard T.M. Vernooij, Johannes A. Lenstra[1], Kai Wang and Leroy E. Hood.

Division of Biology, California Institute of Technology, Pasadena, CA 91125 and [1]Institute of Infectious Diseases and Immunology, Faculty of Veterinary Medicine, University of Utrecht, The Netherlands.

SUMMARY.

Murine T cells carry either of 2 antigen receptors on their surface: αß or γδ T-cell receptors (TCR). In the γδ population subsets exist with tissue specific localization, with invariant variable regions and with specificity for stress antigens. The genomic organization of the complete TCR γ locus is reported here. It spans 205 kb, verifies the clustered organization of the locus and the predicted order and orientation of the individual γ gene clusters. Thus, the Cγ1 cluster is the 5' most cluster and contains Vγ2, 3, 4 and 5. The organization of this cluster is similar to the human TCR γ locus organization. The Cγ2 cluster is in opposite transcriptional orientations relative to the other γ genes. The Cγ4 cluster is small and represents the 3' most cluster. The genomic organization of the Cγ4 gene reveals that the hinge region is encoded by multiple exons. This is similar to the organization of the human Cγ2 gene and unlike the organization of the other mouse and human Cγ gene. The sequence of 2 previously uncharacterized potential γ chain enhancer elements is also described.

INTRODUCTION.

The role of the immune system is to protect the body from foreign entities, by neutralizing and eliminating infections and abnormal cell growth. Among the cell types involved in this system, T cells are indispensable for the proper functioning of the immune system. They provide antigen-specific T-cell help for immunoglobulin secreting cells (B cells) and destroy cells infected with pathogens. Two types of antigen receptors can be found on the surface of T cells: αß or γδ T cell receptors (TCRs). These receptors are generated by somatic rearrangement of V, D (in ß and δ loci only) and J gene segments that are separated in the germline DNA, during T cell differentiation (Kronenberg *et al.*, 1986).

The γδ cells are an unusual T cell population. Certain subsets localize exclusively or preferentially to certain epithelia. Some of these subsets express nearly invariant antigen receptors, indicating that they may recognize tissue specific self antigens. The γδ populations in the skin and gut have been reported to recognize self antigens which are presumably induced by stress (Havran *et al.*, 1991; Eghtesady and Kronenberg, 1992).

The TCR γ chain and the gene segments that encode it are unusual in several respects. Some of the γ chains found in the periphery are homogeneous in their V-J junctions and are expressed exclusively with Vδ1 (Asarnow *et al.*, 1988; Havran

et al., 1989; Itohara *et al.*, 1990; Nandi and Allison, 1991). In the germline of BALB/c mice, the γ gene segments are organized in 4 clusters, each with a single J(oining) gene segment and C(onstant) gene. One of these appears to be non functional and one may be in a transcriptional orientation which is opposite to that of the other 3 clusters (Hayday *et al.*, 1985; Garman *et al.*, 1986; Traunecker *et al.*, 1986; Iwamoto *et al.*, 1986). The locus has a limited number of V(ariable) genes, there are only 7. The Cγ1 cluster contains 4 Vγ gene segments (V2-5) and a T cell specific enhancer (Garman *et al.*, 1986; Pelkonen *et al.*, 1987; Spencer *et al.*, 1991; Kappes *et al.*, 1991). The Cγ3 cluster has only a single V, Vγ1.3 (Traunecker *et al.*, 1986). The Cγ2 cluster only undergoes rearrangement with Vγ1.2, which is located just 5' to the Cγ4 cluster. The Cγ4 cluster also contains only a single Vγ gene segment, Vγ1.1 (Hayday *et al.*, 1985; Iwamoto *et al.*, 1986).

The Cγ4, Cγ2 and Cγ3 clusters are within 150 kb of one another, based on the results of field inversion gel electrophoresis experiments. The Cγ1 cluster has not been linked to the other clusters (Woolf *et al.*, 1988). Molecular cloning experiments have not yet linked the 4 Vγ gene segments in the Cγ1 cluster, nor have they been linked to the Jγ1 gene segment or the Cγ1 gene. Similarly, Vγ1.2 has not been linked to the Cγ2 cluster by molecular cloning (Raulet, 1989).

Infrequent rearrangement between a V from one γ cluster and a J from another cluster does occur. This process deletes the intervening J gene segments and C genes. This characteristic has been used to determine the relative order of the clusters. These analyses established that the 5' to 3' order of the γ chain clusters is as follows Cγ1 - Cγ3 - Cγ2 - Cγ4. The transcription orientation of the Cγ3 and Cγ2 clusters was unknown. However, the orientation of the Cγ2 gene and Jγ2 gene segment was predicted based on the transcriptional orientation of the Vγ1.2 gene (Hayday *et al.*, 1985, Iwamoto *et al.*, 1986).

This report shows the complete genomic structure of the murine TCR γ locus. The clusters are organized as predicted. The Cγ1 cluster in the locus shows distinct homology to the human TCR γ locus, both in organization and in sequence (Huck *et al.*, 1988). Two sequence elements with homology to the γ enhancer are described. The organization of the Cγ4 gene is reported as well. It contains 2 exons that together encode the hinge region of the Cγ4 chain, a situation which is similar to the organization of the human Cγ2 gene.

MATERIALS AND METHODS.

Amplification reactions. Amplifications (Saiki *et al.*, 1988) were performed in 20-150 μl reactions with 1 μM each primer, 2.5 units AmpliTaq/100 μl (Perkin Elmer-Cetus), 50 mM Tris, pH 8.5, 50 mM NaCl, 2.5 mM $MgCl_2$ and 2 mM DTT. Amplification conditions were typically: 30 sec. 93° C, 40 sec. 60° C and 30-120 sec. 72° C for 25-30 cycles. The products were analyzed on agarose gels (1-2%) and isolated in low melt agarose (0.8-1.2%, SeaPlaque, FMC) when appropriate. Primers (5' to 3') for amplifying the enhancer fragments were: 95: AGATCTACTTCCTGATACTCC and 96: AAGCTTAGGGATATTTTGTCTC. The Cγ4 primers were, C54: CAAACGCACTGACTCAGACT, C43: GGAAACAGAATCTCTTGGTCAAC, 133: AAGATAGCAGGACTCTCCTG, 135: ATTGCATGAAAGGAAGAAAGG, 136: ACAACTATTAATGTTAATCCCA and 137: TGGGATTAACATTAATAGTTGT.

Probes. DNA fragments were generated by amplifying template DNA (plasmid or cosmid) with appropriate primers in 20 ul. Typically, 1 ul was reamplified in 150 μl, isopropanol precipitated and separated on a low melt agarose gel. Approximately 50 ng (2-5 μl of the excised band) was used for random priming (Feinberg and Vogelstein, 1983).

Field inversion gel electrophoresis. High molecular weight DNA was isolated from BALB/c liver and embedded in low-melt agarose (InCert, FMC) as described (Popko *et al.*, 1987)

digested with Ksp I and separated on a 1% agarose gel by field inversion gel electrophoresis (FIGE). The gel was blotted onto a nylon membrane (Zetaprobe, BioRad), which was then hybridized with various γ probes (described above).

Genomic cosmid library screening. A BALB/c mouse cosmid library was constructed in vector pWE15A (K. Wang *et al.* in preparation). The γ probes used in the library screening were Vγ1, Cγ2 and a probe derived from the 1 kb γ enhancer (γ1E) described by Spencer *et al.* (1991). All hybridizations were performed according to established protocols (Ausubel, *et al.*, 1991). The cosmid clones were rescreened until pure.

Restriction mapping. The isolated cosmids were characterized with probes for the known Vγ and Cγ genes, as well as with the γ1E probe. DNA was prepared from 1.5 ml cultures according to the alkaline lysis method, from 500 ml cultures via CsCl purification, or from 150 ml cultures via a scaled up alkaline lysis miniprep procedure, followed by a PEG precipitation (Ausubel *et al.*, 1991). Isolated cosmid DNA was digested with Eco RI, or any of a variety of rare cutting enzymes. The DNA samples were separated on agarose gels (0.25 - 1.5%), and transferred to nylon membranes (Zetaprobe). The map of the locus with the location of individual gene segments was constructed from hybridization

data. The distance between gene segments and cosmid ends was determined by PCR amplification.

Cosmid walking. Ends of appropriate cosmids were sequenced with ^{32}P labeled vector primers by cycle sequencing with a commercial kit (BRL), following the manufacturer's instructions. The reactions yielded 180-300 nt. readable sequence. Some cosmid ends were amplified prior to sequencing, in which case the amplified fragment was separated on low melt agarose as described and sequenced *in gel* with the Sequenase kit (USB, Kretz *et al.*, 1989). The obtained sequences were analyzed with the GCG DNA analysis program (Devereux *et al.*, 1984, Altschul *et al.*, 1990, Pearson and Lipman, 1988). Cosmid end-sequences which were not part of known repeats were amplified and used as probes.

RESULTS.

Size of the locus.

The field inversion gel electrophoresis data show that most of the locus is located on a 360 kb Ksp I fragment (figure 1) except Vγ1.1, Jγ4 and Cγ4. These 3 gene segments are on an adjacent 100 kb fragment (figure 1 and below). A previous study (Woolf *et al.*, 1988) located the Cγ2 and Cγ4 clusters on a single 90 kb Sal I fragment, adjacent to a 60 kb Vγ1.3-Cγ3 containing fragment. The Cγ1 cluster (with Vγ2-5) was located on a 45 kb Sal I fragment, which the study did not link to the other fragments.

Yac cloning.

A YAC library of C57Bl/6 DNA (S. Tilghman, Princeton) was screened by PCR amplification with primers for Vγ1 and Cγ1. One pool of positive clones was isolated. Hybridization with various probes identified 1 positive clone. It contained Vγ5, 4, 3 and 2, Cγ2 and the γ specific enhancer (Spencer *et al.*, 1991 and see below). FIGE analysis showed that the YAC clone extended only 10-15 kb 3' to Cγ1 (data not shown).

Cosmid mapping.

A BALB/c cosmid library (K. Wang *et al.*, in preparation) of 15 genome equivalents was screened with a Vγ1 probe, a Cγ3 probe and a probe for the γ specific enhancer (see below). On cosmid DNA blots, these probes detect all Vγ1 members,

all Cγ members (at a wash stringency above 0.2 x SSC at 65^0 C) and 3 enhancer-like fragments (see below). Fifty five positive cosmids were isolated. These were analyzed for the presence of mouse TCR γ gene segments by hybridization with various γ probes. Representative cosmids were also analyzed for the presence of Sal I, Sma I and Cla I sites. Based on the hybridization data, 20 cosmid clones were placed on the map as shown in figure 2.

Description of the murine γ locus.

Figure 2 shows the structure of the entire mouse γ locus as deduced from these clones. It spans 205 kb of DNA, with the V-J-C clusters taking up about 100 kb of DNA (figure 2).

A. Coding regions in the clusters.

1). Cγ1 cluster. The Vγ2, 3, 4 and 5 gene segments are located 5' to Cγ1. They are located within a 10 kb region, 17 kb 5' to Jγ1. The order and distance between the Vγ gene segments from 5' to 3' are: Vγ5 -6.6 kb- Vγ2 -5.8 kb- Vγ4 -1.2 kb- Vγ3. The enhancer sequence, γ1E (Spencer *et al.*, 1991) is located 2.5 kb 3' to the polyadenylation site of Cγ1, as determined by PCR amplification with enhancer and Cγ specific primers (data not shown; Spencer *et al.*, 1991; Kappes *et al.*, 1991). The known gene segments in the cluster (Vγ5 to γ1E) span 41 kb (figure 2).

2). Cγ3 cluster. This cluster was isolated previously on overlapping lambda clones (Traunecker *et al.*, 1986). It contains Vγ1.3, Jγ3 and Cγ3. The distance between Vγ1.3 and Jγ3 was reported as 7.8 kb. Our study finds this distance to be 13 kb. The enhancer-like element in this cluster (γ3E, see below) is located approximately 7.5 kb 3' to the polyadenylation site of Cγ3, as determined by PCR amplification (data not shown). The entire cluster (Vγ1.3 to γ3E) spans 29 kb and is located on a 50 kb Sal I fragment (figure 2).

3) Cγ2 cluster. The Cγ2 gene has only been found expressed with Vγ1.2 (Raulet, 1989), which is located upstream of Cγ4 (figure 2, Iwamoto *et al.*, 1986). The transcriptional orientation of the Cγ2 gene and the Jγ2 and Vγ1.2 gene segments is opposite to that of the other γ gene segments, including Cγ4 (figure 2). The distance from Vγ1.2 to Jγ2 is 22 kb. No cosmid was isolated which contained all 3 elements on it. The enhancer-like element in this cluster (γ2E) is located 3.7 kb 3' to the polyadenylation site of Cγ2, as determined by PCR amplification (data not shown). The entire cluster spans 36 kb (Vγ1.2 through γ2E).

4). Cγ4 cluster. Vγ1.1 is located 2.2 kb upstream of Jγ4, which is 1.8 kb 5' to Cγ4 (Traunecker *et al.*, 1986). The gene segments of the Cγ4 cluster (Vγ1.1, Jγ4 and Cγ4) span

13 kb, which makes this a particularly small immunoglobulin-like V-J-C cluster. Vγ1.2 lies only 4 kb 5' to Vγ1.1, in opposite transcriptional orientation (Hayday *et al.*, 1985; figure 2). The Cγ4 and Cγ2 clusters are located on a single 85-90 kb Sal I fragment.

Genomic organization of the Cγ4 gene segment.

The N terminal part of the hinge region of the mouse Cγ1, 2 and 3 and human Cγ1 chains are 10 to 15 amino acids in length and are encoded by single exons. The human Cγ2 gene is polymorphic for the number of hinge encoding exons. There are 2 or 3 exons, which encode 32 and 48 amino acids, respectively (Littman *et al.*, 1987).

The genomic organization of the mouse Cγ4 gene was not previously known. Since the cDNA sequence predicts a hinge of 33 amino acids, experiments were performed to determine its genomic structure. Figure 3 shows a schematic of the Cγ4 cDNA structure and the location of the amplification primers. Primers 45 and 43 amplify a 300 bp fragment, consistent with a single exon for this domain. Amplification of cosmid DNA with primers 45 and 137 generated a 5 kb DNA fragment, indicating the presence of probably 1 intron between these primer sites. Primers 136 and 135 are from the start and the end, respectively, of the hinge region. Together with a 3' primer (#133), they amplify fragments of 2.8 kb and 1.95 kb, respectively (figure 3). The results

from figure 5 show that the protein coding sequence of Cγ4 spans nearly 8 kb, which is much larger than the 2.5 kb of protein coding sequence of the Cγ1, 2 and 3 genes (Hayday *et al.*, 1985; Garman *et al.*, 1986). If the N-terminal part of the Cγ4 hinge is encoded by 2 exons (see Discussion), then the introns in this gene are 4.6, 0.9 and 1.8 kb in size (figure 3, bottom).

B. Non coding sequences.

Three γ enhancer-like elements.

The γ enhancer probe (Spencer *et al.*, 1991) hybridizes to 3 bands on a genomic blot (figure 4). The 13.5 kb Eco RI fragment contains the described functional enhancer (γ1E) fragment (Spencer *et al.*, 1991; Kappes *et al.*, 1991). Hybridizations of cosmid blots with this probe localized these enhancer like elements 3' to Cγ1 (γ1E), Cγ2 (γ2E, 8.5 kb) and Cγ3 (γ3E, 9.0 kb; figure 2). The γ2E and γ3E crosshybridizing fragments were amplified from cosmids with primers designed from the published γ1E sequence and sequenced (figure 5). They are over 96% identical to γ1E, the functional enhancer.

Linking the clusters.

Cosmid γ54-3 links the Cγ1 and Cγ3 cosmid clusters. The distance between the 3' end of cosmid γ72-2 and 5' end of cosmid γ6 is only 150 bp. The sequence of these ends and the

intervening DNA is colinear with the 3' part of a L1 repeat (data not shown). The linkage correlates with the FIGE data, which identified a 60 kb Sal I fragment containing Vγ1.3 and Cγ3 (data not shown; Woolf *et al.*, 1988). The cosmids identify this as a 50 kb fragment, with the 5' Sal I site located just 3' to γ1E in the Cγ1 cluster (figure 2).

The Cγ3 and Cγ2 cosmid clusters overlap by a short distance, as shown by the overlap of cosmids γ91 and γ37-2 (figure 2). An endprobe from the 5' end of cosmid γ37-2 hybridizes to the end of cosmid γ91 (data not shown). The overlap is less than 6 kb.

Cosmid γ3 and most Cγ4 containing cosmids overlap (figure 2). The overlap contains part of a L1 repeat, flanked by sequences homologous to a mouse retrovirus related sequence (Schmidt *et al.*, 1985). Hybridization with endprobes of cosmids γ3 and γ63 confirmed the overlap (data not shown), as did the presence of the Sma I and Cla I sites in both of these cosmids (figure 2).

DISCUSSION.

Described in this report is the cloning of the complete mouse TCR γ locus, the first mouse TCR locus to be completely cloned. Also reported are the sequences of 2 potential enhancers, the genomic organization of the murine γ locus and the genomic organization of the Cγ4 gene. This information helps to understand the evolution of the γ locus and points to functionally important features of the γ locus.

A cosmid map for the murine TCR γ locus.

The map of the locus shows a nonrandom distribution of cosmid clones. All bonafide cosmids that were isolated from the Cγ2, Cγ4 and 3'end of the Cγ3 regions are shown, except that 3 cosmids were isolated that are identical to cos γ37-2. More cosmids were isolated from the Cγ1 and 5'end of the Cγ3 regions than shown. This nonrandom distribution may be due to preferential digestion of certain Sau3A I sites in the Cγ2/4 regions. This may have resulted in few fragments of approximately 40 kb after partial Sau 3AI (the optimal size for cosmid inserts which allows efficient packaging when ligated into the 8.5 kb vector).

The cosmid map predicts Sal I fragments which are in close agreement with the FIGE map (Woolf *et al.*, 1988). The Cγ2 and Cγ4 clusters are both located on a 85-90 kb Sal I fragment (90 kb according to the FIGE data). The Cγ3 cluster

is present on a 50 kb Sal I fragment (60 kb according to the FIGE data). However, these 2 Sal I fragments are not adjacent to one another as Woolf *et al.* (1988) described. They are separated by a 25 kb Sal I fragment. Woolf *et al.* (1988) measured the partial Sal I fragment which contains the Cγ3, 2 and 4 clusters as 150 kb. The cosmid shows that this distance is 175 kb, which is in the range of the FIGE data. Our own FIGE data show this fragment to be 180 kb (data not shown). The Cγ1 cluster is on a 45 kb Sal I fragment. The 5' Sal I site of this fragment is not present in any of the cosmids. It is presumably located just 5' to cosmid γ5.

The isolated cosmids contain all known γ gene segments, 1.2 kb of DNA 5' to Vγ5 (on cosmid γ5) and approximately 12 kb 3' to Cγ4 (on cosmid γ84; see figure 2).

The organization of the murine γ locus.

A. The V and J gene segments and C genes are organized in clusters.

The known sequence elements of the murine γ locus span approx. 205 kb and consists of 4 V-J-C clusters. Figure 2 shows that the order of the clusters is Cγ1 (with Vγ5, 2, 4 and 3, Jγ1 and Cγ1), Cγ3 (with Vγ1.3, Jγ3 and Cγ3), Cγ2 (with Cγ2, Jγ2 and Vγ1.2) Cγ2 (with Vγ1.1, Jγ4 and Cγ4). Thus, the map confirms the previously predicted order and orientation of the clusters (Raulet, 1989, Traunecker *et*

al., 1986; Iwamoto *et al.*, 1986). The Jγ2 gene segment and Cγ2 gene are in opposite orientation relative to most other γ gene segments (figure 2). This was predicted since Vγ1.2 (which is the only V gene to rearrange to Jγ2) is in an orientation opposite to that of Vγ1.1, Jγ4 and Cγ4 (Hayday *et al.*, 1985; Traunecker *et al.*, 1986). The orientation of the Cγ3 cluster was unknown until recently. In a T cell lymphoma derived from SCID mice, both chromosomes had undergone rearrangements involving Vγ genes of the Cγ1 cluster and Jγ3. These cells had lost both Cγ1 gene, indicating that the Cγ1 and Cγ3 clusters are in the same orientation (Schuler *et al.*, 1991).

The clustered organization of the murine γ locus is unlike that of the other T cell receptor loci and most of the immunoglobulin loci (Lai *et al.*, 1989; Wang *et al.*, in preparation). In this respect it only bears resemblance to the mouse Ig lambda light chain and the shark Ig heavy chain loci (Blomberg and Tonegawa, 1982, Kokubu *et al.*, 1988). The clusters take up about half of the γ locus, approximately 120 kb, and the 2 expressed mouse γ regions (γ1 and γ2/4) together encompass only 90 kb.

The size of the locus can be considered small among the imunoglobulin-like loci in the mouse. The mouse α and ß TCR loci are much larger (approx. 1000 and 750 kb resp., Wang *et al.*, in preparation) and so are the immunoglobulin heavy chain (500 kb or larger) and the kappa light chain loci

(estimated to be larger than 1 Mb, Lai *et al.*, 1989). The mouse TCR γ locus is only one quarter larger than the human γ locus (160 kb, Lefranc *et al.*, 1989).

The cosmid map (figure 2) identifies the only 7 Vγ, 4 Jγ gene segments and 4 Cγ gene that have been identified to date. Additional gene segments might exist in the BALB/c locus, but (1) they do not crosshybridize to the described gene segments and (2) they must rarely rearrange to a Jγ gene segment, as they have not been identified in T cell lines. Nonfunctional Vγ genes may exist, similar to V-like sequences which have been found in the human γ locus (Lefranc *et al.*, 1986; Forster *et al.*, 1987; Huck *et al.*, 1987; Chen *et al.*, 1988).

B. Potential enhancers near Cγ2 and Cγ3.

A functional γ enhancer (γ1E) is located 3' to Cγ1 (Spencer *et al.*, 1991; Kappes *et al.*, 1991). It hybridizes to 2 other fragments, which are located 3' to Cγ2 and Cγ3 (γ2E and γ3E; see figures 2 and 4). These 2 elements are over 96% identical to γ1E (figure 5). Several short sequence elements in γ1E may have a function in enhancer activity. Within the EcoRV - PvuII fragment described by Kappes *et al.* (1991) as having enhancer activity, γ2E has only 2 differences with γ1E in these sequence elements, and γ3E only 3 differences (figure 5). Whether these differences are sufficient to impede enhancer activity can only be determined by

functional studies. It is unlikely that functional γδ cells would ever use γ3E as an enhancer for Cγ3, since Cγ3 and Jγ3 are both nonfunctional gene segments (Hayday *et al.*, 1985; Traunecker *et al.*, 1986). This gene cluster is likely to be nonessential as it is absent from several mouse strains (Iwamoto *et al.*, 1986; Klotz *et al.*, 1989).

The γ1E probe did not hybridize to a DNA fragment in the Cγ4 cluster. This cluster may have an enhancer which is dissimilar to the 3 described enhancers.

Organization inside clusters.

The distances between the J gene segments and C genes are identical in the Cγ1, 2 and 3 clusters: 3.8 kb (Hayday *et al.*, 1985; Traunecker *et al.*, 1986). In the Cγ4 cluster, the J-C distance is only 2.2 kb (Traunecker *et al.*, 1986).

In the Cγ1 cluster, the distance between the mouse Jγ1 gene segment and the closest Vγ gene segment (Vγ3) is 17 kb, very similar to that in the human locus (16 kb). In all other mouse TCR loci, this distance is much larger (Wang *et al.*, in preparation). In the Cγ2, γ3 and γ4 clusters, this V-J distance is 22, 13 and 2.2 kb, respectively (figure 2, Iwamoto *et al.*, 1986). The relatively large V-J distance in the Cγ2 cluster correlates with the apparent insertion of a retrovirus related sequence (Schmidt *et al.*, 1985) which is not present in the other V-J regions (data not shown).

The order of the Vγ gene segments in the Cγ1 cluster (3-4-2-

5) reflects their order of activation in ontogeny. The Vγ3, 4 and 5 genes are also expressed in a tissue specific manner, whereas Vγ2 and the Vγ1 gene segments are not (Raulet, 1989; Allison and Havran, 1991).

In the Cγ3 cluster, the distance from Vγ1.3 to Jγ3 has been reported as 7.8 kb (Traunecker *et al.* 1986). This is at variance with our results (13 kb), and may be due to a strain polymorphism.

The close spacing of Vγ1.1 and Jγ4 (2.2 kb) in the Cγ4 cluster does not appear to prevent occasional rearrangement of Jγ4 to other Vγ segments (Vγ5, Pelkonen *et al.*, 1986). The Vγ1.1, Jγ4 and Cγ4 gene segments span only 13 kb (figure 2), which is small for an immunoglobulin like cluster. Only the 10 kb immunoglobulin heavy chain clusters (V-D-J-C) in the shark genome are smaller (Kokubu *et al.*, 1988).

The sequences of the J, C and γ enhancer-like elements in the Cγ1, 2 and 3 clusters are very similar. The genomic organization of these elements in the clusters is also similar (Hayday *et al.*, 1985; Traunecker *et al.*, 1986; Garman *et al.*, 1986; figure 2). This similarity in sequence and organization suggest that recent duplications gave rise to these γ gene clusters.

Genomic organization of the murine Cγ4 gene.

All mammals analyzed to date have at least 2 C γ region genes. The constant regions encoding these γ chains are very

similar, except in the hinge region (Takeuchi *et al.*, 1992 and references therein). The Cγ1 gene in human and the Cγ1, 2 and 3 genes in mouse have a single exon encoding the N terminal portion of the hinge (Raulet, 1989). The human Cγ2 chain is polymorphic for the number of hinge encoding exons (2 or 3 exons; Littman *et al.*, 1987). Like the human Cγ2 chain, the hinge region of the mouse Cγ4 chain is longer, different in sequence from the hinge of the other mouse Cγ chains (Iwamoto *et al.*, 1986) and encoded by at least 2 exons (figure 5). Since all human and mouse Cγ hinge exons encode 10-16 amino acids (Raulet, 1989), it is possible that there are only 2 hinge exons in the Cγ4 gene of BALB/c mice. Some mouse strains express a Cγ4 chain which is even larger than the described Vγ1.1-J4-Cγ4 chain of BALB/c mice (Cron *et al.*, 1990). This could be due to the presence of a longer hinge region in the Cγ4 chain of these mouse strains, similar to the situation in the human Cγ2 gene. Thus, it is possible that the number of exons which encode the hinge region in Cγ4 is polymorphic in mice, similar to the human Cγ2 chain.

The different C gene may be functionally different due to the different hinge regions (Raulet, 1989). Several morphological characteristics of human Cγ1 expressing cells are different from those of Cγ2 expressing cells (Grossi *et al.*, 1989). This may indicate that there are functional differences between γδ cells with different Cγ hinges. It

would then appear likely that either the particular Cγ gene selected for expression and its surrounding sequence elements determine the characteristics of the cell, or that another genetic element exists which selects for a particular functional characteristic and a particular Cγ gene.

Comparison between murine and human TCR γ loci.

Even though the mouse and human TCR γ loci appear quite different in organization, they have some features in common. Most notably, several similarities exist between the mouse Cγ1 cluster and human locus (figure 6; Pelkonen *et al.*, 1986).

(1) The V-J distance is nearly identical in both loci (17 and 16 kb).

(2) Both loci are compact in the regions which contains the V gene segments, with a spacing of less than 10 kb between V gene segments.

(3) The human locus has only 4 Vγ subfamilies with functional V genes (Vγ1, 9, 10 and 11; see figure 6). The 4 mouse Vγ genes in the Cγ1 cluster are all functional (figure 2).

(4) When comparing the V gene sequences (Huck *et al.*, 1988), it was noted that the mouse Vγ5 and human Vγ1 family members are homologous (the 5' most Vγ gene family on both loci). Homology also exists between murine Vγ2/Vγ1 gene segments

and human Vγ11 and 10 (Figure 6; Huck *et al.*, 1988; chapter 3). The Cγ1, 2 and 3 genes in mouse and both human Cγ gene also have distinct homology (Raulet, 1989). Based on the genomic organization and length of the hinge regions, there appear to be 2 types of Cγ chains: mouse Cγ1, 2, 3 and human Cγ1 vs. mouse Cγ4 and human Cγ2.

The major difference between the mouse and human γ loci is the location of the V genes. In human, all V gene segments are in the 5' half of the locus and both Cγ genes are in the 3' end. In the mouse, each of the 4 C's has at least 1 Vγ gene linked to it.

In summary, even though certain aspects of the loci are different, the overall organization of the mouse Cγ1 cluster and human TCR γ locus are remarkably similar.

Implications of mouse-human comparisons.

The mouse has 3 γδ populations which show specific localization in epithelia in the skin, tongue, female reproductive organs and gut. The Vγ gene segments uniquely or preferentially used in these subsets are Vγ3, 4 and 5 respectively (Raulet, 1989). At least 2 of these 3 mouse γδ populations (skin and gut) recognize tissue specific self components (Havran *et al.*, 1991; Allison and Havran, 1991; Eghtesady and Kronenberg, 1992). In humans, there is no skin specific γδ subset. Thus it not surprising that humans have no mouse Vγ3 homolog (expressed uniquely in mouse skin). The

absence of a mouse Vγ4 homolog in humans may foreshadow that humans have no γδ subset in the tongue and female reproductive organ (not investigated). Humans do have γδ T cells in the gut epithelium, but these cells show no preferential V gene segment usage (Ullrich *et al.*, 1990; Jarry *et al.*, 1990; Deusch *et al.*, 1991). However, the human Vγ1 gene segments are the homologs of mouse Vγ5 (expressed in mouse intestinal epithelium; Huck *et al.*, 1988). The function of the human Vγ1 gene segments may have shifted from gut specific antigen recognition to recognition of other antigens.

No tissue specific localization exists for Vγ1 or 2 expressing γδ T cells. A large proportion of mouse Vγ1.1-Cγ4 expressing T cells appear to be able to respond to mycobacterial antigens, irrespective of the V-J junctions (Happ *et al.*, 1989). The human Vγ9 gene can likewise be stimulated in a polyclonal fashion by several antigens, in a reaction which appears to be a superantigen response (Rust *et al.*, 1990; Band *et al.*, 1991; Ohmen *et al.*, 1991; Pfeffer *et al.*, 1991 Sturm *et al.*, 1991). Possibly, the human Vγ9 gene segment may be a functional counterpart for the mouse Vγ1.1 gene segment.

Evolution of the mouse γ locus.

The Cγ1, 2 and 3 genes are very homologous to one another, as are the Jγ1, 2 and 3 segments, the Vγ1 family members

(Hayday *et al.*, 1985; Garman *et al.*, 1986; Traunecker *et al.*, 1986) and the 3 described γ enhancer-like elements (figure 5). The Jγ4 and Cγ4 gene shows distinct but lower homology with the other *J* gene segments and C genes as well (66%, Iwamoto *et al.*, 1986; Traunecker *et al.*, 1986). And Vγ2 shows some homology to the Vγ1 members. These homologous gene segments and their location in the locus suggest that they arose by duplication. During evolution, a primordial Vγ-Jγ-Cγ locus may have duplicated to give rise to the Vγ2-Cγ1 and Vγ1.1-Cγ4 clusters. Subsequent duplication of Vγ1.1 and Jγ1-Cγ1 (including the enhancer) in the mouse lineage may have created a new Vγ1-Jγ-Cγ cluster, which, upon complete duplication, may have given rise to the Vγ1.2-Cγ2 and Vγ1.3-Cγ3 clusters. The other Vγ genes (Vγ3, 4 and 5) were perhaps not involved in the gene duplication, or they may have been deleted later in all clusters except the Cγ1 cluster. Subsequent events may have changed the details of the organization of the clusters, such as the V-J distances. Nevertheless, the events outlined above would explain the overall organization of the murine γ locus. However, more complicated explanations can be proposed, including gene conversion which makes (distantly) related gene segments similar. The genomic organization of the TCR γ loci of other mammals may shed additional light on the evolution of the locus.

We have isolated the physical DNA of the entire mouse TCR γ locus. Sequence analysis of this locus and the human locus may give a better understanding of how the mouse and human loci have evolved, as well as identify important elements in the locus, including promoter and additional enhancer (near Cγ4?) elements.

Acknowledgements.

We thank Dr. Masaaki Toda for help in the isolation of the YAC clone.

REFERENCES.

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

Altschul, S.F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403-410.

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

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1991). Current Protocols in Molecular Biology.

Band, H., Porcelli, S. A., Panchamoorthy, G., McLean, J., Morita, C. T., Ishikawa, S., Modlin, R. L., and Brenner, M. B. (1991). Antigens and antigen-presenting molecules for gamma/delta T cells. Curr. T. Micr. *173*, 229-234.

Blomberg, B., and Tonegawa, S. (1982). Organization and expression of mouse lambda light chain immunoglobulin genes. In: Gene Regulation. B. O'Malley (ed.). Academic Press, New

York.

Chen, C. H., Cihak, J., Losch, U., and Cooper, M. D. (1988). Differential expression of two T cell receptors, TcR1 and TcR2, on chicken lymphocytes. Eur. J. Immunol. *18*, 539-543.

Cron, R. Q., Coligan, J. E., and Bluestone, J. A. (1990). Polymorphisms and diversity of T-cell receptor-gamma proteins expressed in mouse spleen. Immunogenetics *31*, 220-228.

Deusch, K., Luling, F., Reich, K., Classen, M., Wagner, H., and Pfeffer, K. (1991). A major fraction of human intraepithelial lymphocytes simultaneously expresses the gamma/delta T cell receptor, the CD8 accessory molecule and preferentially uses the V-delta-1 gene segment. Eur. J. Immunol. *21*, 1053-1059.

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

Eghtesady, P., and Kronenberg, M. (1992). Intestinal $\gamma\delta$ T lymphocytes are autoreactive for stressed intestinal epithelial cells. Submitted.

Feinberg, A. P., and Vogelstein, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. *132*, 6-13.

Forster, A., Huck, S., Ghanem, N., Lefranc, M. P., and Rabbitts, H. (1987). New subgroups in the human T cell rearranging V-gamma gene locus. Embo. J. *6*, 1945-1950.

Fox, V. L., Strauss, W. M., and Seidman, J. G. (1989). Isolation and restriction map of the V-J interval of human T cell receptor gamma chain locus. Genomics *4*, 445-448.

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

Grossi, C. E., Ciccone, E., Migone, N., Bottino, C., Zarcone, D., Mingari, M. C., Ferrini, S., Tambussi, G., Viale, O., Casorati, G., Millo, R., Moretta, L., and Moretta, A. (1989). Human T cells expressing the gamma/delta T-cell receptor (TcR-1): C-gamma-1- and C-gamma-2-encoded forms of the receptor correlate with distinctive morphology, cytoskeletal organization, and growth characteristics. Proc. Natl. Acad. Sci. USA *86*, 1619-1623.

Happ, M. P., Kubo, R. T., Palmer, E., Born, W. K., and

O'Brien, R. L. (1989). Limited receptor repertoire in a mycobacteria-reactive subset of gamma/delta T lymphocytes. Nature *342*, 696-698.

Havran, W. L., Chien, Y., and Allison, J. P. (1991). Recognition of self antigens by skin-derived T cells with invariant gamma/delta antigen receptor. Science *252*, 1430-1432.

Havran, W. L., Grell, S., Duwe, G., Kimura, J., Wilson, A., Kruisbeek, A. M., O'Brien, R. L., Born, W., Tilelaar, R. E., and Allison, J. P. (1989). Limited diversity of T-cell receptor gamma-chain expression of murine Thy-1+ dendritic epidermal cells revealed by V-gamma-3-specific monoclonal antibody. Proc. Natl. Acad. Sci. USA *86*, 4185-4189.

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 gamma genes. Cell *40*, 259-269.

Huck, S., and Lefranc, M. P. (1987). Rearrangements to the JP1, JP and JP2 segments in the human T-cells rearranging gamma gene (TRG-gamma) locus. FEBS. LETTERS. *224*, 291-296.

Huck, S., Dariavach, P., and LeFranc, M. P. (1988). Variable

region genes in the human T-cell rearranging gamma (TRG) locus: V-J junction and homology with the mouse genes. Embo. J. *7*, 719-726.

Itohara, S., Farr, A. G., Lafaille, J. J., Bonneville, M., Takagaki, Y., Werner, H., and 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 gamma genes in C57BL/10 mice. Sequence and expression of new constant and variable region genes. J. Exp. Med. *163*, 1203-1212.

Jarry, A., Cerf-Bensussan, N., Brousse, N., Selz, F., and Guy-Grand, D. (1990). Subsets of CD3+ (T cell receptor alpha/beta or gamma/delta) and CD3- lymphocytes isolated from normal human gut epithelium display phenotypical features different from their counterparts in peripheral blood. Eur. J. Immunol. *20*, 1097-1103.

Kappes, D. J., Browne, C. P., and Tonegawa, S. (1991). Identification of a T-cell-specific enhancer at the locus encoding T-cell antigen receptor gamma chain. Proc. Natl. Acad. Sci. USA *88*, 2204-2208.

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. J. 7, 3413-3422.

Kretz, K. A., Carson, G. S., and O'Brien, J. S. (1989). Direct sequencing in low-melt agarose with sequenase. Nucleic Acids Res. *17*, 5864-5870.

Lai, E., Wilson, R. K., and Hood, L. (1989). Physical maps of the mouse and human immunoglobulin-like loci. Adv. Immunol. *46*, 1-59.

Lefranc, M. P., Forster, A., Bear, R., Stinson, M. A., and Rabbitts, T. H. (1986). Diversity and rearrangement of the human T cell rearranging gamma genes: Nine germ-line variable genes belonging to two subgroups. Cell *45*, 237-246.

Lefranc, M., and Rabbitts, T. H. (1989). The human T-cell receptor gamma (TRG) genes. TIBS *14*, 214-218.

Littman, D. R., Newton, M., Crommie, D., Ang, S-L., Seidman, J. G., Gettner, S. N., and Weiss, A. (1987). Characterization of an expressed CD3 associated Ti-γ chain reveals Cγ domain polymorphism. Nature *326*, 85-88.

Nandi, D., and Alliso, J. P. (1991). Phenotypic analysis and gamma-delta T cell receptor repertoire of murine T cells associated with the vaginal epithelium. J. Immunol. *147*, 1773-1778.

Ohmen, J. D., Barnes, P. F., Uyemura, K., Lu, S., Grisso, C. L., and Modlin, R. L. (1991). The T cell receptors of human gamma/delta T cells reactive to mycobacterium tuberculosis are encoded by specific V genes but diverse V-J junctions. J. Immunol. *147*, 3353-3359.

Pearson, W. R., and Lipman, D. J. (1988). Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA *85*, 2444-2448.

Pelkonen, J., Traunecker, A., and Karjalainen, K. (1987). A new mouse TCR V-gamma gene that shows remarkable evolutionary conservation. Embo. J. *6*, 1941-1944.

Pfeffer, K., Schoel, B., Gulle, H., Moll, H., Kromer, S., Kaufmann, S. H. E., and Wagner, H. (1991). Analysis of

primary T cell responses to intact and fractionated microbail pathogens. Curr. T. Micr. *173*, 173-178.

Popko, B., Puckett, C., Lai, E., Shine, H. D., Readhead, C., Takahashi, N., Hunt III, S., Sidman, R. L., and Hood, L. (1987). Myelin deficient mice: Expression of myelin basic protein and generation of mice with varying levels of myelin. Cell *48*, 713-721.

Raulet, D. H. (1989). The structure, function, and molecular genetics of the gamma/delta T cell receptor. Ann. Rev. Immunol. *7*, 175-207.

Rust, C. J. J., Verreck, F., Vietor, H., and Koning, F. (1990). Specific recognition of staphylococcal enterotoxin A by human T cells bearing receptors with the V-gamma-9 region. Nature *346*, 572-574.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science *239*, 487-491.

Schmidt, M., Wirth, T., Kroeger, B., and Horak, I. (1985). Structure and genomic organization of a new family of murine retrovirus-related DNA sequences (MuRRS). Nucleic Acids Res.

13, 3461-3470.

Schuler, W., Ruetsch, N. R., Amsler, M., and Bosma, M. (1991). Coding joint formation of endogenous T cell receptor genes in lymphoid cells from scid mice: unusual P-nucleotide additions in VJ-coding joints. Eur. J. Immunol. *21*, 589-596.

Spencer, D. M., Hsiang, Y., Goldman, J. P., and Raulet, D. H. (1991). Identification of a T-cell-specific transcriptional enhancer located 3' of C-gamma-1 in the murine T-cell receptor gamma locus. Proc. Natl. Acad. Sci. USA *88*, 800-804.

Sturm, E., Braakman, E., Fisch, P., Sondel, P.M., and Bolhuis, R. L. H. (1991). Daudi cell specificity correlates with the use of a V-gamma-9 V-delta-2 encoded TCR-gamma-delta. Curr. T. Micr. *173*, 183-188.

Takeuchi, N., Ishiguro, N., and Shinagawa, M. (1992). Molecular cloning and sequence analysis of bovine T-cell receptor gamma and delta chain genes. Immunogenetics *35*, 89-96.

Traunecker, A., Oliveri, F., Allen, N., and Karjalainen, K. (1986). Normal T cell development is possible without functional gamma chain. Embo. J. *5*, 1589-1593.

Ullrich, R., Schieferdecker, H. L., Ziegler, K., Riecken, E. O., and Zeitz, M. (1990). Gamma/delta cells in the human intestine express surface markers of activation and are preferentially located in the epithelium. Cell. Immunol. *128*, 619-627.

Woolf, T., Lai, E., Kronenberg, M., and Hood, L. (1988). Mapping genomic organization by field inversion and two-dimensional gel electrophoresis: Application to the murine T-cell receptor gamma gene family. Nucleic Acids Res. *16*, 3863-3875.

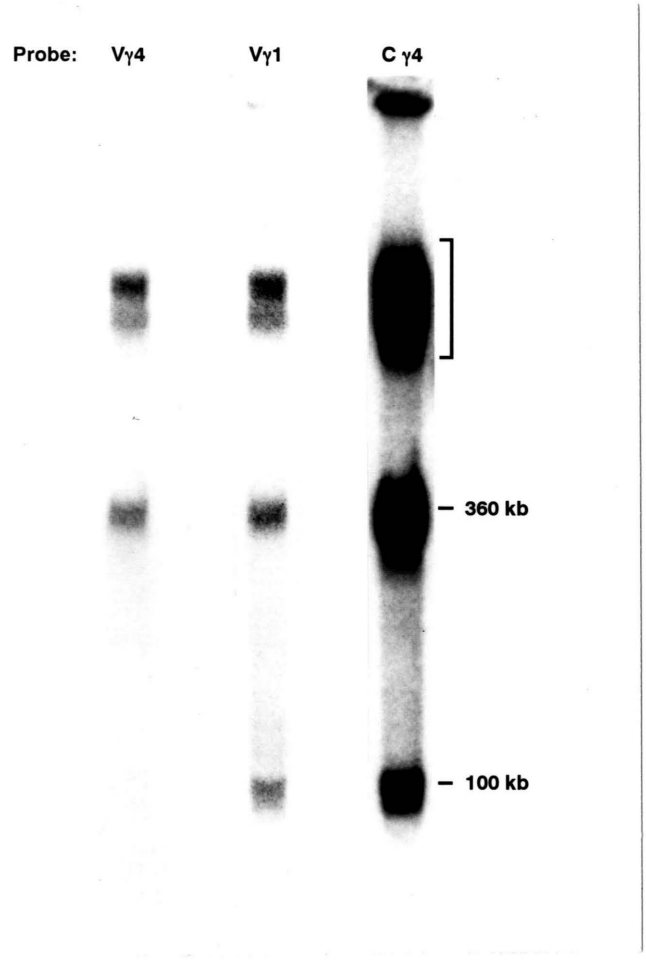


Figure 1. Mapping of the murine TCR γ locus by field inversion gel electrophoresis. A Ksp I blot of BALB/c DNA was hybridized to a Vγ4, Vγ1.2 and a Cγ4 probe. All 3 probes identify the same 360 kb band. The Vγ1.2 and Cγ4 probes also identify a 100 kb fragment. The hybridization signal in the top of the gel corresponds to partially digested DNA.

Figure 2. Map of the murine T cell receptor γ locus. The gene segments and genes are indicated as vertical lines (V and J segments) or as boxes (C and E segments). Individual exons of the V gene segments and C genes are not indicated. The J gene segments were not mapped in this study, but were placed based on previous reports (Hayday *et al.*; 1985, Traunecker *et al.*, 1986; Iwamoto *et al.*, 1986). Nonfunctional gene segments and genes are indicated by ψ. Infrequent restriction enzyme sites are indicated (C= Cla I; K=Ksp I; S= Sal I; Sm= Sma I).

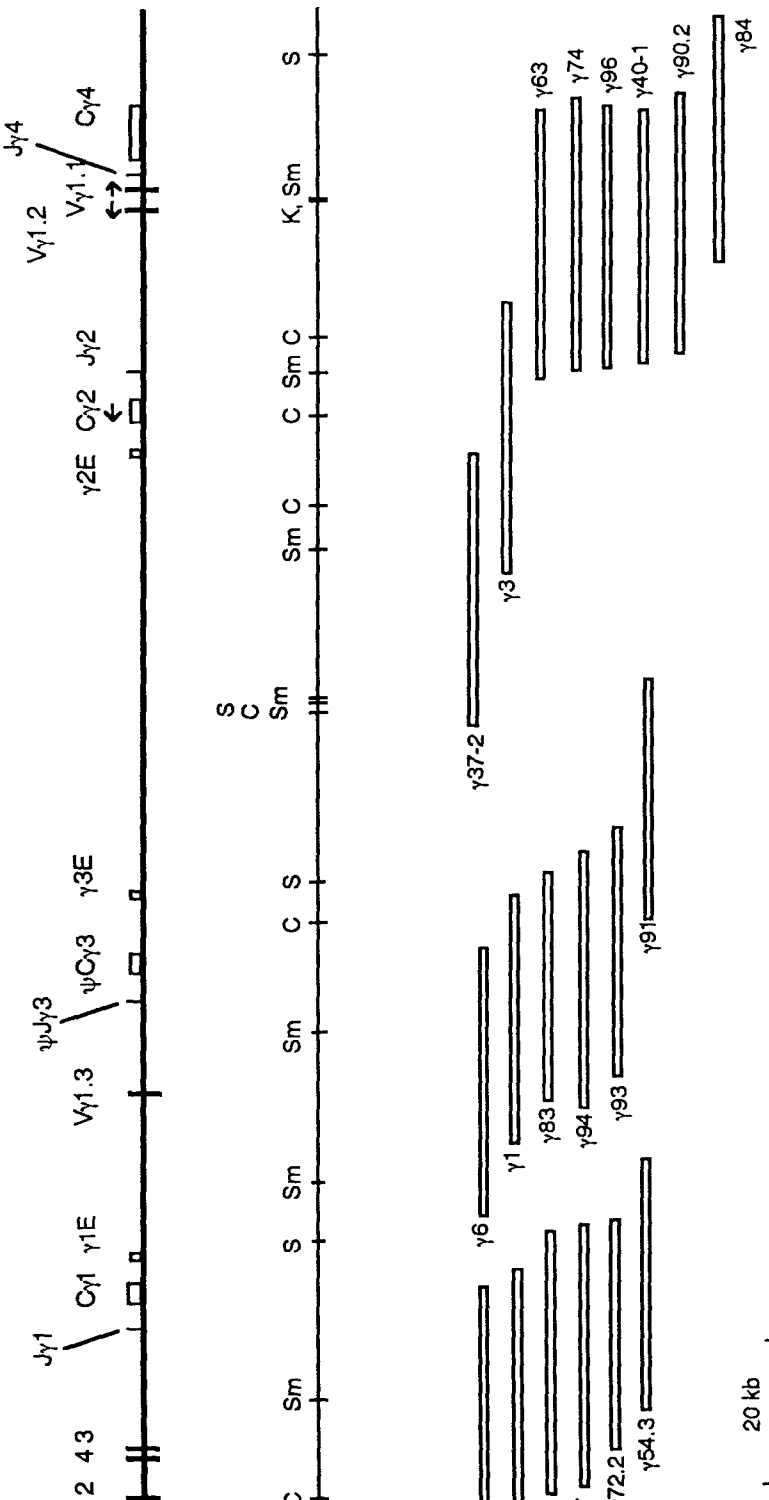
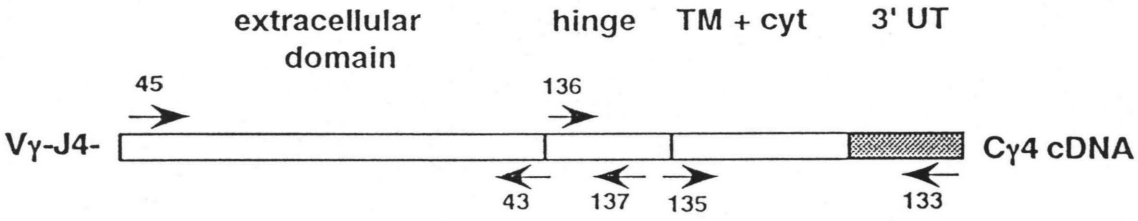
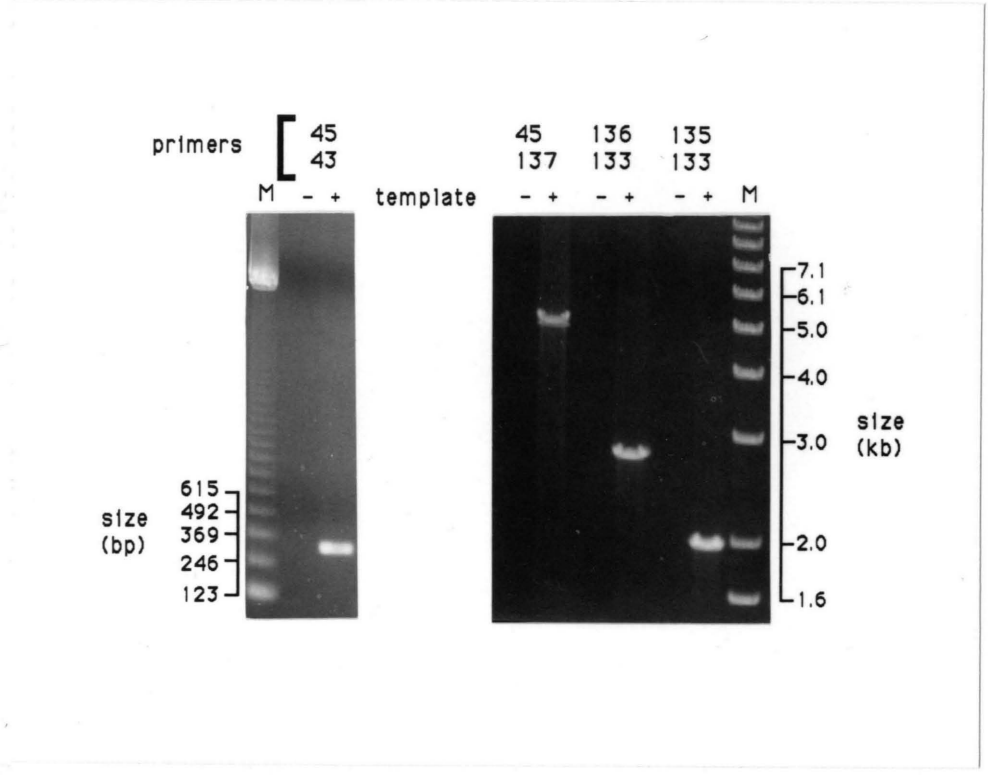

Vγ5 2 4 3
Jγ1
Cγ1 γ1E
Vγ1.3
ψJγ3
ψCγ3 γ3E
γ2E Cγ2 Jγ2
Vγ1.2
Jγ4
Vγ1.1
Cγ4
C Sm S Sm Sm C S S C Sm Sm C C Sm C K, Sm S
γ5 γ4 γ90.1 γ7 γ72.2 γ54.3 γ6 γ1 γ83 γ94 γ93 γ91 γ37-2 γ3 γ63 γ74 γ96 γ40-1 γ90.2 γ84
20 kb

Figure 3. The Cγ4 gene in BALB/c is encoded by at least 4 exons and spans over 8 kb. **A.** Schematic representation of the cDNA sequence of Cγ4 and the putative exon boundaries (based on comparison with the Cγ1, 2 and 3 sequences; Iwamoto *et al.*, 1986; Raulet, 1989); TM: transmembrane region, cyt: cytoplasmic domain, 3' UT: 3' untranslated region, shaded box). The primers which were used for amplification are shown as arrows. **B.** Result of amplification of cosmid γ84 DNA with various combinations of primers. **C.** Proposed genomic organization of the BALB/c Cγ4 gene as deduced from the amplification data (see text). The exons are indicated as boxes. The intron sizes are indicated below the figure. The size of the 3' untranslated region of Cγ4 is not known.

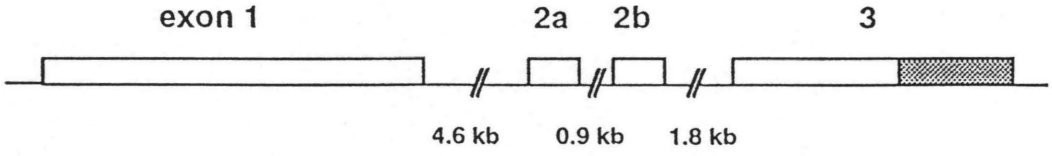
A.


extracellular domain
hinge
TM + cyt
3' UT
45
136
Vγ-J4-
Cγ4 cDNA
43
137
135
133

B.


primers
45
43
M
- +
template
45
137
136
133
135
133
- +
- +
- +
M
7.1
6.1
5.0
4.0
3.0
2.0
1.6
size (kb)
size (bp)
615
492
369
246
123

C.


exon 1
2a
2b
3
4.6 kb
0.9 kb
1.8 kb

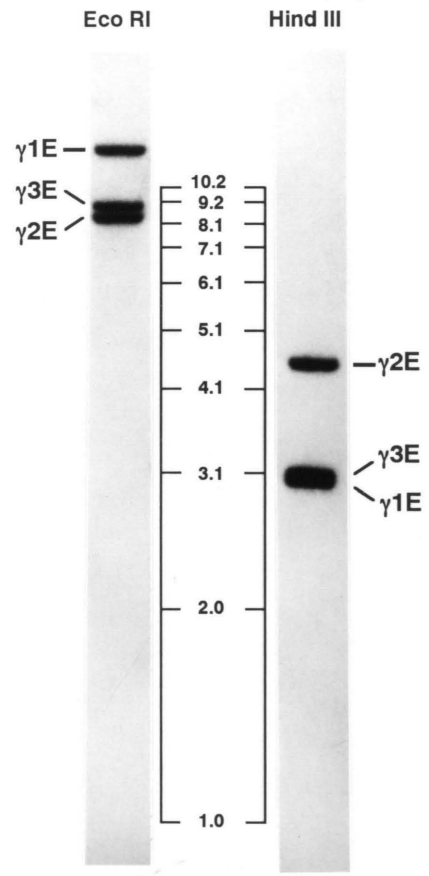
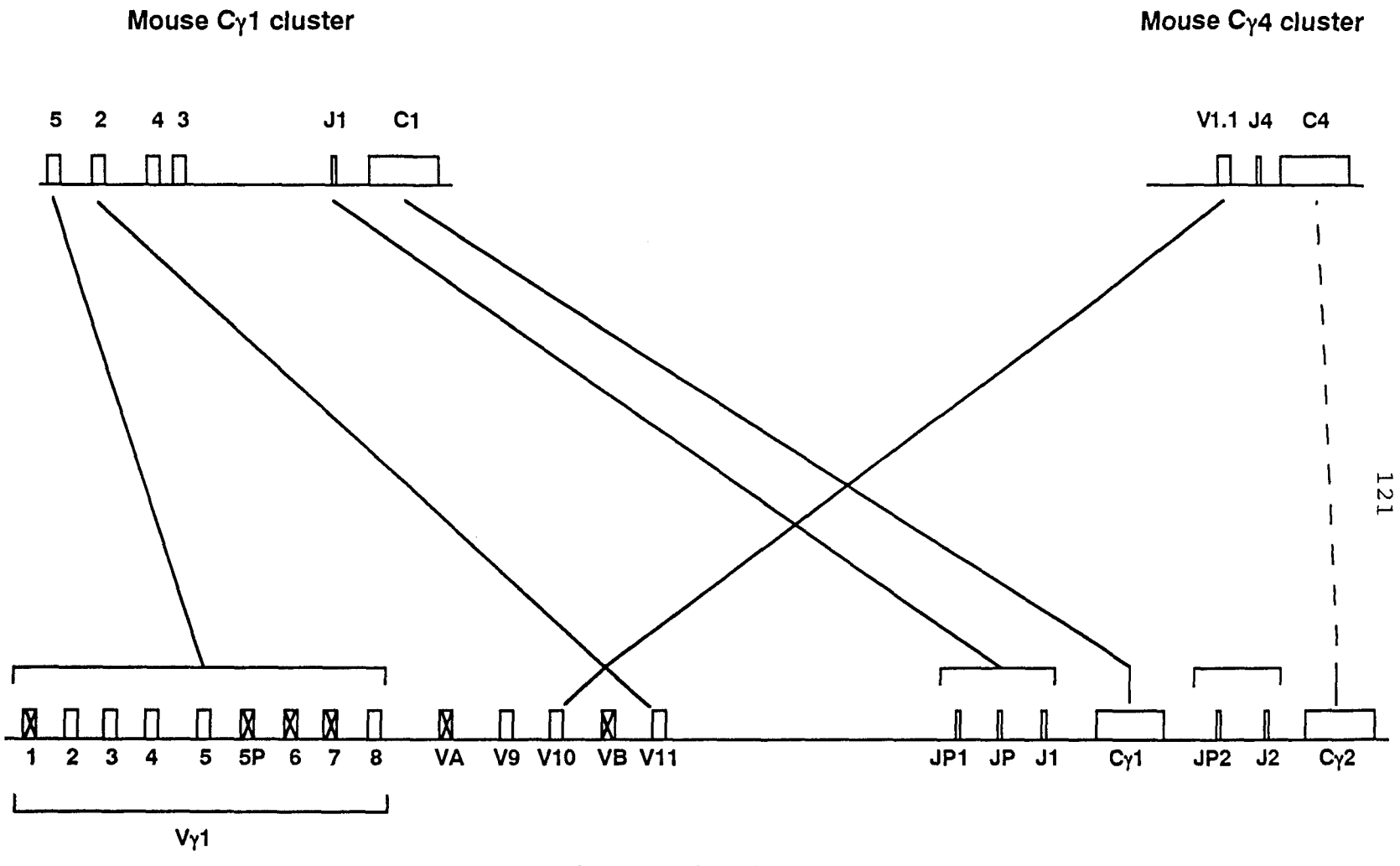


Figure 4. The γ enhancer hybridizes to 3 Eco RI and 3 Hind III fragments in BALB/c genomic DNA. The γ1E probe was hybridized to a genomic blot containing BALB/c DNA.

Figure 5. BALB/c mice have 2 sequences which are very similar to the γ enhancer. The γ1E sequence is from Spencer *et al.* (1991). The γ2E and γ3E segments were amplified from cosmids with primers 95 and 96 and sequenced in low-melt agarose with specific primers. Gaps were introduced into the aligned sequences to maximize identity. Sequences which may be important for enhancer function are boxed (from Spencer *et al.*, 1991).

```
γ1E AGATCTACTT CCTGATACTC CTGAACAGAC TAGACCTCCC CAGGCCCTGT TCCTAGCCTC TAAGCAGAGG CATGGCTATG TCAGCACTAG GAAACAGATG  100
γ2E                                    TC.T. .......... ....G..... ...A....TT .......... .......... ..........
γ3E                                          .......... .NN....... ...N..---. .......... .......... ..........

γ1E CCATGCCTGG ----GAACTG ACAATAGGCA GGAGTCT--- -----TGATA TCATTGGTCA ATTGCAGCAG ACATGTCTCT TGTGGTGTAA AGCCTCATAG  200
γ2E .....A.... CTGG..... .......... .......TGT GTGGC..... .......... .......... .......... .......... ..........
γ3E .......... ----...... .A........ .......--- -------... .......... .......... .......... .......... ..........

γ1E CATCTTCCTG CTGGAGTCCA CCTGGCAGCC CAAAGCCTCA GCAAGTGCTG GGGACAGAGA GATAGAAAGG CTTTGCTGTG GCTCAACTCA GTTGTACTTT  300
γ2E .......... .......... .......... .......... .......... .......... .......... .......... .......... ...T......
γ3E .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........

γ1E CTCAAGCTCC TTTTGGTTAT CTGCTCTGTC TATGTAGTTG CTTCCTGGAA AATGGTTAAA GTCAAACCTC CTCTGTTTG GGCAGCACAA AATGTTGCTA  400
γ2E .......... .......... ..T....... .......... .......... .......... .......... .......... .......... ..........
γ3E .......... .......... .......... ......A... .......... .......... .......... ----------. .......... ..........

γ1E AGGCGTTAGA AGACAACATA GGAGCAGTTA AACCACAGCC AGTTTTTGCT CGCTTTCGAA AGACCACAGC TATTAGACAG AAAACCTCCT GTACTGGAAG  500
γ2E .......... ....T..... ..C....... .......... .......... .......... .......... .......... .......... ..........
γ3E .......... ....T..... ..C....... .......... .......... .A........ .......... .......... .......... ..........

γ1E TGGCCCTGAA ACAGCATCTG TGCTGACAGA CACTGGGCTT GCTATGGTTG CTGCTCATCA TGCTCAGCTG AGCATGGAAT ATGCTTCCCG GAGCTGCGTA  600
γ2E ...T..C... .......... ..A....... .......... .......... ..T....... .......... .......... C......... ..........
γ3E ...T...... .......... .......... .......... .......... .......... .......... .......... C......... ..........

γ1E GACACTGGAC CCAAGAAAGG AACAGAAAAG GAGAGATGTG TCCCCGTATC ATGTTATAAT TCCATCAGAA GTTTTTTCTC TAGCACTTGA TATAAGTGTT  700
γ2E .......... .......... .......... .......... .......... .......C.. .......... .......... .......... ..........
γ3E .......... .......... .......... .......... .......... .......C.. .......... .......... .......... .........A

γ1E CAAGCAGCTT TTGCTTCATT TTCTAAGAAG ATTAAAAAAA GGAACAAAGC TGTGTCCTGA ACAAAGTAGC ATAGGAGAAT GAAGAAGTGA GCAAGAGCAA  800
γ2E ....T..... .......... .......TG. ...T...... .......... .......... .......... .......... .....---... ..........
γ3E ....T..... .......... .......TG. ...C...... .......... ....G..... .......... .......... .....---... ..........

γ1E TGTGCTATAG GAAGTCACAG AAAAAGATAA CCAAGGACAG C-TGTGCTAC TTGCCATTCA CTGAGGAGCA TCCACCCCAG GAGGAAAATG GTATCAGAGT  900
γ2E .......... ..T....... ...T...... .......... .-........ .......... .......... .......... .......... ..........
γ3E .......... ..T....... ...T...... .......... .T........ .......... .......... .......... .......... ..........

γ1E CCTCTGAGAA AGGC-TAACT GAAAATATTA ACAGAATTA- CAAGTACACA ACAGACTAGC ATCAGAAAAT AAATGTTCAG AAATATGAAC ACAGAGACAA 1000
γ2E .......... ....T..... .......... .........- .......... .......... .......
γ3E .......... ....T..... .....A.... ......A..T ......T... .......... .......

γ1E AATATCCCTA AGCTT
γ2E
γ3E
```

Figure 6. Homology between the mouse Cγ1 and Cγ4 clusters and the human TCR γ locus. Solid lines indicates sequence homology and the dashed line indicates homology in genomic organization. The schematic representation of the human TCR γ locus is from data in Lefranc and Rabbitts,1989 and Fox *et al.*, 1989.


Mouse Cγ1 cluster
Mouse Cγ4 cluster
5 2 4 3 J1 C1
V1.1 J4 C4
1 2 3 4 5 5P 6 7 8 VA V9 V10 VB V11 JP1 JP J1 Cγ1 JP2 J2 Cγ2
Vγ1
Human TCR γ locus
121

Chapter III.

STRUCTURE AND EVOLUTION OF T CELL RECEPTOR GAMMA GENE SEGMENTS.

SUMMARY.

The T cells in the mammalian body are essential components of the response to pathogens. They recognize components of the infections via the T cell receptor (TCR). The γδ T cell receptor (TCR) is composed of 2 chains. The γ chain is encoded by genes and gene segments located in the TCR γ locus. These genes and gene segments have been conserved during evolution. All examined species have at least 2 Cγ genes. The Cγ genes of mouse, man, sheep and cow are very similar, except in the hinge, or connector region. This hinge region is variable in length and in sequence. The Vγ gene segments encode proteins which are quite variable in sequence, but have conserved features which presumably allow the V domains to fold similarly. Evolutionarily, the C genes may have arisen from 2 ancestral Cγ genes, which were present before the lineages of primates, rodents and ruminants diverged.

The V gene elements appear to have been less well conserved. Only 3 Vγ gene subfamilies in mouse and man show distinct sequence similarity. The other V genes do not have counterparts in the other species, indicating that they have been lost in one but not the other lineage.

INTRODUCTION.

T cells play a crucial role in the immune system. They provide help in the induction of antibody production and cause lysis of cells infected with viruses and other pathogens. These T cells are antigen specific, in that they can only be activated by a specific structure. The receptor which recognizes this antigen structure is termed the T cell receptor (TCR). The TCRs are heterodimers which are expressed on the surface of the T cells. Two types of TCRs exist: one is composed of an α and a ß chain (the αß TCR), and the other of a γ and δ chain (the $\gamma\delta$ TCR).

The TCR gene segments which encode these chains are located in large loci. There are generally many V(ariable) gene segments, 1 or 2 C(onstant) genes and 1, or 2 types of short gene segments (J, or D and J). In a characteristic way, a functional TCR α or γ gene is generated by DNA rearrangement which brings a V gene segment together with a J gene segment. In the ß and δ TCR loci, at least 2 rearrangements must occur in order to generate a functional TCR gene. Here, a V, a D and a J gene segment need to be combined into a single contiguous V gene, by DNA rearrangement (Kronenberg *et al.*, 1986; Davis and Bjorkman, 1988). Transcription of the composite V gene and the C region exons followed by RNA splicing generates mRNA which encodes a γ chain.

Comparisons of the α and ß protein sequences showed that the C domains (Cα and Cß) have similarity to the IgC regions in

the occurrence and position of various amino acids (Williams and Barclay, 1988). The Vα and Vß regions show similarity to the Ig V region. These domains have been postulated to fold in a similar manner (Chothia *et al.*, 1988; Williams and Barclay, 1988).

This paper compares Cγ and Vγ protein and gene sequences of various mammals. The structure of the C regions is well conserved, with the exception of a single subdomain. The V regions are far less conserved, except for some residues which are probably important for proper folding of this domain. The sequence comparisons also show that during evolution, certain species lost TCR γ gene elements, or acquired additional γ elements by gene duplication or gene conversion.

MATERIAL AND METHODS.

Gene sequences. The gene sequences were obtained from Genbank or entered manually into a VAX computer. All further manipulations were performed with the GCG programs (Devereux *et al.*, 1984).

Alignments. The nucleotide sequences were translated into protein sequences, and aligned using the GCG LINEUP program. Alignments were based on maximal identity between the amino acid residues. Gaps were introduced when necessary for optimal alignment. The sequences were then pairwise scored for percent identity. Each mismatch with a gap was scored as a difference, unless the difference at the C-terminus was due to one sequence being shorter than the other.

The nucleotide sequence alignment was based on the protein sequence alignment. Percent identity was scored between all pairwise combinations as for the proteins.

RESULTS AND DISCUSSION.

Table 1 lists the γ genes and gene segments, the references in which they were reported and the name under which they are listed in the figures. The 3' ends of some Vγ gene segments are not known, as they were only reported as cDNA sequences. In these cases, the nucleotides which were potentially N-region nucleotides were omitted from the sequences before alignment. The human Cγ2 gene sequence used in the alignment and comparisons includes all 3 copies of exon 2 (Littman *et al.*, 1987).

I. *Constant regions.*

A. Protein structure.

Figure 1 shows an alignment of the protein sequences of all known Cγ chains. The consensus shown on the top line identifies all positions where 6 or more of the 10 residues at that position contain the same amino acid. The putative ß strands are underlined (based on comparisons with sequences reported in Lesk and Chothia (1982); Williams and Barclay (1988)).

All mammals examined to date have a minimum of 2 Cγ genes. It is clear that all chains share significant homology, with most differences located in the C-terminal half of the C regions. The C regions consist of an extracellular, immunoglobulin constant region-like domain of 110 amino acids, a transmembrane region (TM) of 20-22 amino

Table 1. Sequences and references.

	name used	reference
mouse		
Cγ1	Mgc1	Hayday *et al.* (1985)
Cγ2	Mgc2	Hayday *et al.* (1985)
Cγ4	Mgc4	Iwamoto *et al.* (1986)
Vγ1.1	Mgv1.1	Hayday *et al.* (1985)
Vγ1.2	Mgv1.2	Hayday *et al.* (1985)
Vγ1.3	Mgv1.3	Traunecker *et al.* (1986)
Vγ2	Mgv2	Garman *et al.* (1986)
Vγ3	Mgv3	Garman *et al.* (1986)
Vγ4	Mgv4	Garman *et al.* (1986)
Vγ5	Mgv5	Pelkonen *et al.* (1987)
human		
Cγ1	Hgc1	Lefranc *et al.* (1986a)
Cγ2	Hgc2	Lefranc *et al.* (1986a)
		Littman *et al.* (1987)
Vγ1 family		Lefranc *et al.* (1986b)
	Hgv1-2 to Hgv1-8	
Vγ9	Hgv9	Forster *et al.* (1987)
		Huck *et al.* (1988)
Vγ10	Hgv10	Forster *et al.* (1987)
		Huck *et al.* (1988)
Vγ11	Hgv11	Huck *et al.* (1988)

Table 1. (cont.) Sequences and references.

	name used	reference
sheep		
Cγ1	Sgc1	Hein *et al.* (1990)
Cγ2	Sgc2	Hein *et al.* (1990)
cow		
Cγ1	Bgc1	Takeuchi *et al.* (1992)
Cγ2	Bgc2	Takeuchi *et al.* (1992)
Cγ3	Bgc3	Takeuchi *et al.* (1992)

acids and a cytoplasmic region, generally of 12 residues. Only 6 amino acids are present in bovine Cγ1 (Hein *et al.*, 1990; Littman *et al.*, 1987). A hinge or connector region of variable length connects the IgC like and TM domains.

Extracellular Ig like domain.

The extracellular immunoglobulin constant region (IgC) like domains are similar in length in all mammalian Cγ proteins (figure 1). These domains contain 2 cysteine residues characteristic of Ig constant and V region like domains, at positions 32 and 88 (figure 1). The putative disulfide bond between these 2 residues spans 56 amino acids in the Cγ proteins of all examined species. Obvious regions of homology are present in the sequences as shown in figure 1. The IgC domain has been postulated to make 7 ß strands,

termed A to G (Lesk and Chothia, 1982; Williams and Barclay, 1988). The sequences which encode strands A, B (which contains the Cys at position 32) and E have been most conserved. In the other strands, conserved and nonconserved amino acid changes are evident. The extracellular IgC like domain of the Cγ proteins appears to be well conserved.

The transmembrane and cytoplasmic regions.

The TM regions of the Cγ proteins are well conserved too (boxed in figure 1). This 20-22 amino acid region contains a hydrophilic and potentially charged Lys residue in the mouse, human and one sheep protein. This residue has been postulated to interact with acidic residues in the TM region of CD3 chains (Raulet, 1989). Interestingly, this charged residue is not present in the TM domain of the sheep Cγ2 and bovine Cγ proteins. It will be interesting to determine if these TcR heterodimers interact with CD3 components in a way similar to those of γ chains which contain a positively charged residue in the TM region.

The cytoplasmic region of the Cγ region is 12 amino acids in length in all but 1 protein (bovine Cγ1). It appears that a mutation in bovine Cγ1 truncated this chain by 6 amino acids (see the nucleotide sequence in figure 2: the codon in bovine Cγ1 is TGA, one nucleotide different from TGT in bovine Cγ2). All chains contain a transmembrane anchor in this region (2 Arg residues in most chains) and a Cys

(except bovine Cγ1). The TM and cytoplasmic domains are well conserved in addition to the extracellular domain among the Cγ proteins of the examined species. This suggests that these regions have important functions which depend on specific protein structures.

The hinge region.

The most obvious difference between the various Cγ chains is in the hinge region, which connects the extracellular IgC like and TM domains. The length of the hinge is quite variable and ranges from 24 amino acids in mouse Cγ2 to 79 residues in bovine Cγ1. Most of the variability resides in the N-terminal end of the hinge. The C-terminal end of all hinges is well conserved. This region contains the conserved Cys residue (position 173 in figure 1) which is thought to be involved in the intrachain disulfide bond with the TcR δ chain. Only human Cγ2 lacks this Cys residue and as expected, Cγ2 containing γ chains are not covalently linked to a δ chain (Littman *et al.*, 1987). The sheep Cγ1 and 2 and bovine Cγ1 and 2 proteins all have 2 additional Cys residues in the hinge regions, spaced less than 20 amino acids apart.

The hinge regions are thought to allow flexibility due to the presence of Pro residues (Kronenberg *et al.*, 1986). All chains but mouse Cγ2 have prolines in this domain. Sheep Cγ1 and all bovine γ chains have 4 or more Pro residues. These

hinge regions could thus allow considerable flexibility in the Cγ chains, with the exception of the mouse Cγ2 chain. The hinge of mouse Cγ2 is the shortest of the Cγ hinges (24 residues).

Several of the longer hinge regions show evidence of duplicated stretches of amino acids. This has been shown for the human Cγ2 chain, where 2 or 3 exons (each encoding 16 amino acids: residues 111-127, 128-142, 143-178 in figure 1) can be included in the hinge. These exons are highly similar to one another and presumably arose by triplication of a single exon. The human Cγ1 gene has only a single copy of this very homologous exon (Littman *et al.*, 1987). The hinge of the sheep and bovine Cγ chains contain repeated motifs of the short sequence TTEPP. Bovine Cγ1 has 4 related motifs, sheep Cγ1 has 3, bovine Cγ2 has 2 and sheep Cγ2 and bovine Cγ3 have only a single copy. This could be due to a variable number of highly homologous exons in the respective genes (Takeuchi *et al.*, 1992).

The N-terminal portion of the hinge regions of the human and mouse Cγ chains display very limited homology with the hinges of the ruminant Cγ proteins (figure 1). The N-terminal region of the hinge of the mouse Cγ4 chain is unusual in that it shows no obvious resemblance to any of the other hinge regions. Only weak homology to the hinge of the human Cγ genes is apparent.

The function of the divergent hinge exons is not known. It

is interesting that all mammals examined have at least 2 types of Cγ chains: one with a short hinge and one with a longer hinge. This length heterogeneity may have a functional significance, or may simply be a reflection of the fact that the sequence is not important (other than the conserved C-terminal end with the Cys residue): it may simply act as a spacer between the IgC like domain and the TM region and allow the dimeric TcR to be flexible.

B. Gene structure.

The genomic organizations of the mouse Cγ1, 2, 3 genes and the human Cγ1 gene are very similar. Exon 1 encodes the extracellular IgC like domain, exon 2 encodes the part of the hinge which includes the Cys residue (10-16 amino acids) and exon 3 encodes the rest of the hinge (20 amino acids), the TM and the cytoplasmic domains and the 3' untranslated region. The human Cγ2 gene organization is similar to the human Cγ1 gene, except that the number of exon 2 segments is polymorphic in the human Cγ2 gene: 2 or 3 copies are present (Littman *et al.*, 1987). The structure of the mouse Cγ4 gene has not been determined in detail, but PCR analysis has shown that there are at least 2 exons which encode the hinge region segments (chapter 2). This genomic organization with multiple hinge exons exist in the human Cγ2 gene and possibly the sheep and bovine Cγ genes as well (Takeuchi *et al.*, 1992).

C. Evolutionary implications.

Figure 2 shows an alignment of the nucleotide sequences of the Cγ genes. The alignment was based on the protein alignment of figure 1.

The similarity between the Cγ chains is obvious from figure 1 and 2, and is also represented in tables 2 and 3. Since the hinge regions are the most divergent between the various Cγ genes, the similarities in tables 2 and 3 are indicated in 2 ways: as overall identity (total) and as identity excluding the amino acids from positions 111 to 181 in figure 1 (indicated as 'minus hinge' in tables 2 and 3). The following discussion will primarily use the latter comparisons.

Two or more Cγ genes in each species.

The mouse Cγ1 and 2 gene sequences are highly similar to one another (95%) and also to the mouse Cγ3 gene (>94%, data not shown). Cγ3 is probably a defective gene, as it has a splice site defect (Hayday *et al.*, 1985; Garman *et al.*, 1986). It seems likely that these 3 mouse Cγ genes arose by a recent duplication, or have undergone recent gene conversion events. The other mouse Cγ gene (Cγ4) is considerably more divergent (approx. 75% outside the hinge regions).

The human Cγ genes are also very similar to one another, especially when only a single copy of exon 2 in the human Cγ2 gene is considered (amino acids 143-179 in Hgc2, figure

1; Littman *et al.*, 1987; Lefranc *et al.*, 1986). Their close similarity indicates that they are the result of a recent duplication or gene conversion event.

The described bovine Cγ1 and 2 chains are also very similar, except for the length of the hinge (98% identity outside the hinge region). These 2 chains may even be encoded by allelles, containing different numbers of hinge encoding exons (Takeuchi *et al.*, 1992). The third bovine Cγ gene (Cγ3) is more divergent from bovine Cγ1 and 2 (82% identity outside the hinge region).

The 2 sheep Cγ genes are relatively divergent, as they show only limited homolgy to one another (85% outside the hinge). In the 3 mammals other than human, 2 divergent types of Cγ genes exist (mouse Cγ1, 2, 3 - mouse Cγ4, sheep Cγ1 -sheep Cγ2 and bovine Cγ1, 2 - bovine Cγ3). In human, the 2 Cγ genes are very similar, except for the hinge region.

The Cγ genes from all species show similarity.

From the comparison of the bovine and sheep genes (figure 2 and table 4), it is clear that the sheep Cγ1 and bovine Cγ1 and 2 genes shared a common ancestor (95% identity outside the hinge). A high degree of similarity between sheep and cow orthologous genes is not surprising, as the sheep and cow diverged from a common ancestor only 20 million years (Myr) ago (Irwin *et al.*, 1991). The presence of a second Cγ gene in both species which are only 85% similar (outside the

hinge region) is surprising. Either the 2 chains present in the common ancestor diverged at very different rates after sheep and cow diverged, or more than 2 chains were present in this ancestor. In sheep one Cγ gene may have been lost and in cow another.

The mouse, human and ruminant lineages arose approximately 80 Myr ago. The exact time of divergence has not been conclusively established (Pesole *et al.*, 1991). The data in table 2 show that the mouse Cγ1 and 2 are most homologous to the human Cγ1 and 2 genes (80% outside the hinge) indicating that these 2 species may be more closely related than to the ruminants. There appears to be no counterpart to mouse Cγ4 in human, sheep or cow. As all studied mammals have at least 2 (generally divergent) Cγ genes, it is likely that the common ancestor to all these species already had at least 2 Cγ genes. In all species but mouse, the Cγ4 ancestral gene may have been replaced by a duplicate copy of the other gene.

The similarity between the mouse and human Cγ genes (except Cγ4) outside the hinge regions (80%) is very similar to that between the mouse and human Cα and Cδ coding regions (76% and 79% respectively; Koop *et al.*, 1992).

Proposed evolution of the TCR γ Constant genes.

Figure 3 shows a diagram of how the TCR Cγ genes in the 4 studied mammals may have evolved from a common ancestor. The

diagram is based on the assumptions that the ancestral locus had 2 Cγ genes, that the bovine Cγ1 and 2 genes are allelles (Takeuchi *et al.*, 1992), and that cow and sheep have no other Cγ genes.

In this proposed evolution scheme, the ancestral Cγ gene which gave rise to the mouse Cγ4 gene was lost in the other 3 mammals. All other Cγ genes present in these species are more similar to the mouse Cγ1, 2 and 3 genes than to the mouse Cγ4 gene, especially when the hinge region is not included in the comparison (table 3).

The timing of gene duplications (i.e., recent vs. ancient) can not be determined. An ancient duplication followed by a recent gene conversion involving the entire coding sequence will look like a recent duplication. There is some indication that gene conversion did indeed occur in the mouse TCR γ locus, namely in the enhancer-like elements 3' to the mouse Cγ genes (chapter 2; data not shown). As the Cγ genes in all 4 mammals show similarity, they too may have undergone gene conversion to make the genes within each species similar, rather than by gene duplication.

The mouse Cγ1, 2 and 3 genes probably arose by recent gene duplications (or underwent a recent gene conversion), as they are over 94% homologous to one another.

The human Cγ1 and 2 genes are also very homologous (98% outside the hinge), indicating that they too arose by (recent) gene duplication. Alternatively, the ancestral gene

indicated in figure 3 as C4 may not have been lost, but gene converted to make both human Cγ genes similar.

The ruminant evolution is somewhat more complex. Sheep Cγ1 and bovine Cγ1 are 95% similar (outside the hinge), indicating that they arose from the same ancestral Cγ gene. The sheep Cγ2 gene is 85% similar to the sheep Cγ1 gene and to the bovine Cγ1/2 and Cγ3 genes, and less than 80% similar to any of the mouse or human Cγ genes. The bovine Cγ3 gene is also more similar to the other ruminant Cγ genes than to the mouse and human genes. This indicates that all ruminant Cγ genes have a common ancestor. Since the sheep Cγ2 and bovine Cγ3 genes are only 85% similar (vs. 95% between sheep and bovine Cγ1), they probably had separate ancestors before the divergence of sheep and cow. Thus, the ruminant ancestor must have had 3 Cγ genes, one giving rise to sheep and bovine Cγ1, one giving rise to sheep Cγ2 and one giving rise to bovine Cγ3. There is no indication that ancestral gene 'C4' has been retained in the ruminants. But again, the loss of the ancestral 'C4' gene is equivalent to it being gene converted by the C1 ancestral gene.

This proposed evolutionary pathway shows some interesting features. Despite loss of sequence information (Cγ4) in all but one lineage, all species still have at least 2 Cγ genes. In each species, the 2 (or more) Cγ genes have hinges of different length. Second, gene duplication (and/or gene conversion) events have occurred in all lineages. The

plasticity of the TCR γ loci is further indicated by the different organization of these loci in mouse and human (chapter 2). The human locus has an organization similar to most other TCR and Ig loci: a single cluster exists with the Vs upstream, the 2 C genes all downstream and the J elements upstream of the C elements. The mouse locus on the other hand consists of 4 clusters, each with one or more Vs, one J and one C gene.

It will be interesting to determine the organization of the TCR γ locus of the 2 ruminant species for which γ sequences are known. The organization of these mammals may indicate how the loci have evolved and what the structure of the ancestral locus was.

In summary, all species examined have multiple Cγ genes. The general stucture of these chains has been well conserved, except for the absence of a charged residue in the TM region of most ruminant proteins, the lack of a Cys in the hinge of human Cγ2 and the variability in length and sequence of the N-terminus of all hinge regions. The sequence comparisons indicate that the γ genes have undergone multiple events during evolution, including gene duplication and/or gene conversion.

II. *Variable regions.*

A. Protein structure. The only species for which complete V gene sequences have been reported are mouse and human. Several V gene families exist in these 2 mammals, including mouse Vγ1 (3 members) and human Vγ1 (7-10 members, not all of which are functional; Lefranc *et al.*, 1986b). An alignment of the functional V region amino acid sequence is shown in figure 4. The proposed ß strands (based on alignments with sequences reported in Williams and Barclay (1988) and Chothia *et al.* (1988)), are indicated.

The size of the V regions varies from 102 to 106 residues. From the alignment it is evident that far fewer amino acids residues are conserved in the V regions than in the C regions. Only 20 residues are present in similar positions in 60% or more of these chains. These include the conserved Cys residues (at positions 27 and 103 in the consensus sequence in figure 4). Since all sequences have some type of deletion relative to the consensus, the proposed disulfide bond loops out a somewhat variable number of amino acids: 69 (mouse Vγ3) to 73 residues (3 other Vγ chains). Only the human Vγ11 chain lacks the first Cys residue (figure 4). Several conserved amino acids are present around these conserved Cys residues. They are part of the proposed ß strands B and F, respectively (Chothia *et al.*, 1988; Williams and Barclay, 1988). In ß strand C, 4 amino acids are conserved: IHWY. The latter 2 amino acid are also

conserved in the V regions of α and ß TCR genes (Chothia *et al.*, 1988). All but 1 amino acid residue shown in the consensus lie in the proposed ß strands which are thought to be involved in intra- and inter chain interactions (Chothia *et al.*, 1988; Williams and Barclay, 1988). Their conservation would predict that they have an important role in the structure of the V domain.

The size of the loops between ß strands B and C and C' and D are variable due to deletions/insertions in the sequences (amino acid 35-38 and 64-67). These 2 loops constitute the hypervariable regions (hv1 and hv2) which are thought to be part of the antigen binding site of the TcR. The third hypervariable loop is encoded by the VJ junction of the rearranged gene segments (Chothia *et al.*, 1988; Williams and Barclay, 1988; Davis and Bjorkman, 1988).

B. Evolutionary implications.

The Vγ gene sequence alignment is shown in figure 5. It is apparent that the mouse Vγ1 members are very similar to one another (94-95%, figures 4 and 5, table 3B). They are arranged in the mouse γ locus as single V genes upstream of separate J and C genes (chapter II). Their similarity suggests that they arose by duplication, possibly in tandem with J and C genes (chapter 2). They show some similarity to mouse Vγ2 (65-72%; table 3B). The similarity between the mouse Vγ1 members and the other mouse Vγ genes is not

significant at the amino acid level (21-31%), but limited homology can be detected with Vγ3 and 4 at the nucleotide level (40-45%; table 3B).

The human Vγ1 proteins are clearly very similar to one another (66-91%), but show little similarity to any of the other human V regions (table 3A). More limited similarity among the human Vγ genes exists between Vγ10 and 11 (43% at the amino acid level and 59% at the nucleotide level). Some similarity also exists between human VγB and Vγ10/11, but only at the nucleotide level (data not shown). The VγB gene is a pseudogene and therefore not included in the alignments.

The mouse Vγ5 region shows similarity to the human Vγ1 members (approx. 46 and 65% at the protein and nucleotide level, respectively; table 4A and 4B; Huck *et al.*, 1988). In the human lineage, the Vγ1 ancestral gene duplicated several times to give rise to multiple Vγ1 genes. The number of human Vγ1 members is polymorphic and ranges from 7 to 10 (Ghanem *et al.*, 1989). Several of these are pseudogenes (Lefranc *et al.*, 1986b).

In mouse, Vγ5 expressing cells are located mainly in the gut, in or near the epithelial cell layer (Takagaki *et al.*, 1989; Bandeira *et al.*, 1991). Recent experiments suggest that these T cells respond to stress antigens, produced by

the gut epithelial cells. This in vitro response could be inhibited by antibodies to TL antigens, indicating that these nonclassical class I MHC molecules present the stress antigen(s) to the Vγ5 expressing γδ T cells (Egthesady *et al.*, 1992; Eghtesady and Kronenberg, 1992). No preferential expression of any of the TcR Vγ genes in the human gut has been reported (Ullrich *et al.*, 1990), leaving their function unknown. Whether the human Vγ1 expressing cells are restricted in their response by nonclassical class I MHC molecules, similar to the mouse Vγ5 expressing cells is not known.

The mouse Vγ1 and Vγ2 regions appear to be similar to the human Vγ11 region, and somewhat less similar to human Vγ10 (tables 4A and 4B). The homology between mouse Vγ2 and human Vγ11 is 51% at the amino acid level and 71% at the nucleotide level. The antigen specificity of these gene segments is unknown.

Proposed evolution of the TCR γ Variable genes.

A proposed evolutionary pathway for the human and mouse Vγ gene segments is shown in figure 6. It is partially based on the assumption that V genes which show similarity descended from common ancestors, rather than becoming similar due to convergent evolution. In this diagram, a single V gene (VA in figure 6) gave rise to V5 in the mouse and the V1 family in human.

All other V genes in mouse and human may have descended from a single ancestral V gene (VB in figure 6, top) which existed well before the split of the mouse and human lineages. Before the mouse-human divergence, the VB gene probably duplicated to give rise to 5 genes, labeled C-G in figure 6. These genes gave rise to the other functional mouse and human V genes. Mouse V2 and the Vγ1 members (V1.1, 1.2 and 1.3) evolved from VC, which in the human lineage gave rise to V11. The VC duplication which gave rise to V2 and the V1 members (65-72% similar) may have occurred even before the mouse-human split. The similarity between the mouse and human noncoding region (i.e. with no or minimal selection) in the TCR Cα-Cδ region is approximately 67% (Koop *et al.*, 1992). Thus, coding regions (which are expected to have some level of selection pressure) which are 65-70% similar can be expected to have diverged before the mouse-human lineages diverged.

Figure 6 shows VD to be the ancestor of mouse V4, whereas is has been deleted in the human lineage. Similarly, VE gave rise to mouse V3, but in the human lineage it too was deleted. The human V9 and 10 genes evolved from separate ancestral genes VF and VG (figure 6). These genes were lost in the mouse lineage.

In summary, 2 groups of related V gene segments have been identified in mouse and human. Mouse Vγ1 and 2 are

homologous to human Vγ10 and 11, whereas mouse Vγ5 is homologous to the human Vγ1 subfamily. The proposed evolution of these genes shows a variety of gene duplication and deletion events. It will be of interest to analyse the Vγ genes which are present in other mammals. The presence of homologs of the mouse and human Vγ genes in other mammals may indicate whether the proposed evolution is correct or not. It also will show which Vγ genes are conserved among species (maybe mouse Vγ5) and which are not. This may point to a function of this Vγ which is similar in all species, e.g., recognition of similar antigens, presented by similar MHC molecules.

Speculation.

1. Mouse Vγ3 expressing cells are located exclusively in the skin (Havran *et al.*, 1989). The skin of humans do not show specific localization of γδ cells. This correlates with the absence of a Vγ3 homolog in the human. Perhaps the absence of a mouse Vγ4 homolog in the human correlates with an absence of a γδ T cell subset in the female reproductive organs and tongue, the tissues in which the mouse Vγ4 cells are located (Itohara *et al.*, 1989).

The skin of the cow has abundant γδ T cells. It will be of interest to determine if all these cells use a single Vγ and Vδ gene (Vγ3/Vδ1 in mouse), if all these cells have monomorphic γδ TCR (invariant junctions, as in the mouse),

and whether these cells respond to self antigens expressed by skin cells (keratinocytes in the mouse). A caveat to these speculations is that the mouse γδ cells in the skin are located in the epidermis, whereas in the cow they are located in the dermis, raising the possibility that the cow skin γδ cells may be distinct in that they recognize a self antigen from another cell type.

2. The presence of 2 Cγ chains in all species examined to date raises the question why this is so. One explanation could be that the 2 chains have different functions (or are used in 2 different subsets of γδ T cells). If this is true in humans, then this is most likely due to the difference in the hinge regions, as the 2 Cγ regions in some individuals differ by only a single change outside the hinge regions (Raulet, 1989). Alternatively, the Cγ genes may be functionally equivalent, whereas the Jγ gene segments that are located just 5' to each Cγ gene may determine functional differences. Or there may be a bias in rearrangement, allowing certain V genes to rearrange almost exclusively to one specific J-C gene pair. Selection for such a V gene would then also imply selection for the J-C pair.

The sequence comparisons have shown that the T cell receptor Cγ genes are conserved among the studied mammals. The region in the Cγ genes which is least conserved corresponds to the

hinge region. The Vγ gene segments are far less conserved in sequence. Evidence exists for loss of Vγ gene segments in the human and mouse lineages.

Analysis of the TCR γ loci in other mammals may give clues regarding the importance of phylogenetically conserved sequences, such as the length and composition of the Cγ hinges and the presence of particular V genes in the genome.

REFERENCES.

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 gamma-delta. Proc. Natl. Acad. Sci. USA *88*, 43-47.

Chothia, C., Boswell, D. R., and Lesk, A. M. (1988). The outline structure of the T-cell AlphaBeta receptor. Embo. J. *7*, 3745-3755.

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

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

Forster, A., Huck, S., Ghanem, N., Lefranc, M. P., and Rabbitts, T. H. (1987). New subgroups in the human T cell rearranging V-gamma gene locus. Embo. J. *6*, 1945-1950.

Egthesady, P. and Kronenberg, M. (1992). Intestinal γδ T lymphocytes are autoreactive for stressed intestinal epithelial cells. Submitted.

Eghtesady, P., Panwala, C. Teitell, M., and Kronenberg, M. (1992). Recognition of the thymus leukemia (TL) antigen by intestinal γδ lymphocytes. Submitted.

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

Ghanem, N., Buresi, C., Moisan, J. P., Bensmana, M., Chuchana, P., Huck, S., Lefranc, G., and Lefranc, M. P. (1989). Deletion, insertion, and restriction site polymorphism of the t-cell receptor gamma variable locus in French, Lebanese, Tunisian, and Black African populations. Immunogenetics *30*, 350-360.

Havran, W. L., Poenie, M., Tigelaar, R. E., Tsien, R. Y., and Allison, J. P. (1989). Phenotypic and functional analysis of gamma-delta T cell receptor-positive murine dendritic epidermal clones. J. Immunol. *142*, 1422-1428.

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

Hein, W. R., Dudler, L., Marcuz, A., and Grossberger, D.

(1990). Molecular cloning of sheep T cell receptor gamma and delta constant regions: unusual primary structure of gamma chain hinge segments. Eur. J. Immunol. *20*, 1795-1804.

Huck, S., Dariavach, P., and LeFranc, M. P. (1988). Variable region genes in the human T-cell rearranging gamma (TRG) locus: V-J junction and homology with the mouse genes. Embo. J. *7*, 719-726.

Irwin, D. M., Kocher, T. D., and Wilson, A. C. (1991). Evolution of the cytochrome b gene of mammals. J. Mol. Evol. *32*, 128-144.

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. Proc. Natl. Acad. Sci. USA *86*, 5094-5098.

Iwamoto, A., Rupp, F., Ohashi, P. S., Walker, C. L., Pircher, H., Joho, R., Hengartner, H., and Mak, T. W. (1986). T cell-specific gamma genes in C57BL/10 mice sequence and expression of new constant and variable region genes. J. Exp. Med. *163*, 1203-1212.

Koop, B. F., Wilson, R. K., Wang, K., Vernooij, B., Zaller,

D., Kuo, C. L., Seto, D., Toda, M., and Hood, L. (1992). Organization, structure and function of 95kb of DNA spanning the murine T-cell receptor Cα/Cδ region. Genomics, in press.

Kronenberg, M., Siu, G., Hood, L., and Shastri, N. (1986). The molecular genetics of the T-cell antigen receptor and T-cell antigen recognition. Ann. Rev. Immunol. *4*, 529-591.

LeFranc, M. P., Forster, A., and Rabbitts, T. H. (1986a). Genetic polymorphism and exon changes of the constant regions of the human T-cell rearranging gene gamma. Proc. Natl. Acad. Sci. USA *83*, 9596-9600.

LeFranc, M. P., Forster, A., Bear, R., Stinson, M. A., and Rabbitts, T. H. (1986b). Diversity and rearrangement of the human T cell rearranging gamma genes: Nine germ-line variable genes belonging to two subgroups. Cell *45*, 237-246.

Lesk, A. M., and Chothia, C. (1982). Evolution of proteins formed by beta-sheets. II. The core of the immunoglobulin domains. J. Mol. Biol. *160*, 325-342.

Littman, D. R., Newton, M., Crommie, D., Ang, S., Seidman, J. G., Gettner, S. N., and Weiss, A. (1987). Characterization of an expressed CD3-associated Ti gamma-chain reveals C-gamma domain polymorphism. Nature *326*,

85-88.

Pelkonen, J., Traunecker, A., and Karjalainen, K. (1987). A new mouse TCR V-gamma gene that shows remarkable evolutionary conservation. Embo. J. *6*, 1941-1944.

Pesole, G., Sbisa, E., Mignotte, F., and Saccone, C. (1991). The branching order of mammals: phylogenetic trees inferred from nuclear and mitochondrial molecular data. J. Mol. Evol. *33*, 537-542.

Raulet, D. H. (1989). The structure, function, and molecular genetics of the gamma/delta T cell receptor. Ann. Rev. Immunol. 7, 175-207.

Takagaki, Y., DeCloux, A., Bonneville, M., and Tonegawa, S. (1989). Diversity of gamma-delta T-cell receptors on murine intestinal intraepithelial lymphocytes. Nature *339*, 712-714.

Takeuchi, N., Ishiguro, N., and Shinagawa, M. (1992). Molecular cloning and sequence analysis of bovine T-cell receptor gamma and delta chain genes. Immunogenetics *35*, 89-96.

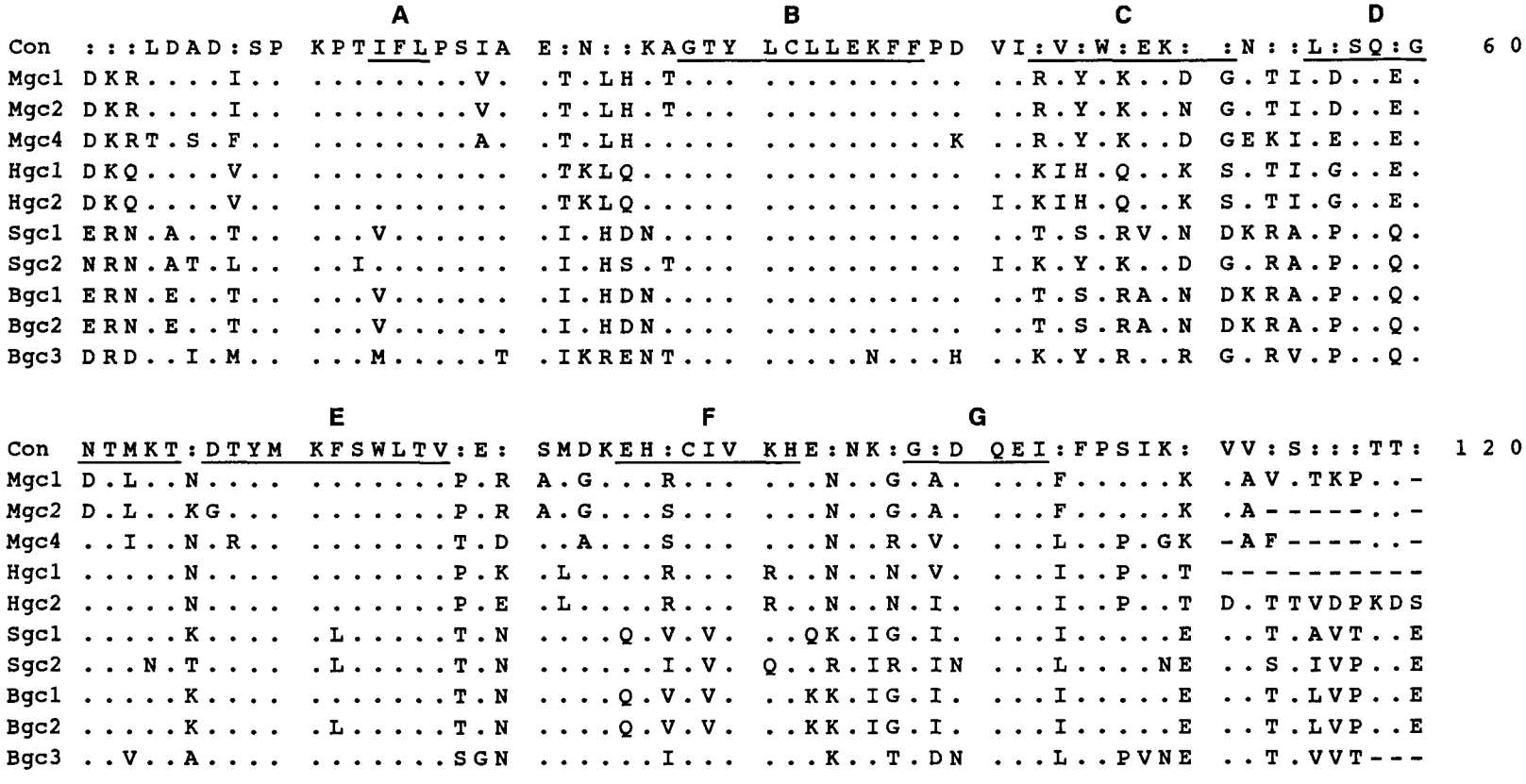
Traunecker, A., Oliveri, F., Allen, N., and Karjalainen, K. (1986). Normal T cell development is possible without

functional gamma chain. Embo. J. *5*, 1589-1593.

Ullrich, R., Schieferdecker, H. L., Ziegler, K., Riecken, E. O., and Zeitz, M. (1990). Gamma/delta cells in the human intestine express surface markers of activation and are preferentially located in the epithelium. Cell. Immunol. *128*, 619-627.

Williams, A. F., and Barclay, A. N. (1988). The immunoglobulin superfamily-domains for cell surface recognition. Ann. Rev. Immunol. *6*, 381-405.

Figure 1. Alignment of TCR γ constant region proteins. A consensus (Con) is shown where 6 or more amino acids are identical. A ":" in the consensus sequence indicates that no consensus was apparent. A "-" indicates a gap which was introduced to optimize the alignment. A "." denotes identity with the consensus. The "*" identifies a stop codon. The putative ß strands are underlined and the transmembrane region is boxed.


A
B
C
D
Con :::LDAD:SP KPTIFLPSIA E:N::KAGTY LCLLEKFFPD VI:V:W:EK: :N::L:SQ:G 60
Mgc1 DKR....I.. ....V. .T.LH.T... ....R.Y.K..D G.TI.D..E.
Mgc2 DKR....I.. ....V. .T.LH.T... ....R.Y.K..N G.TI.D..E.
Mgc4 DKRT.S.F.. ....A. .T.LH..... ....K ..R.Y.K..D GEKI.E..E.
Hgc1 DKQ....V.. .... .TKLQ..... ....KIH.Q..K S.TI.G..E.
Hgc2 DKQ....V.. .... .TKLQ..... .... I.KIH.Q..K S.TI.G..E.
Sgc1 ERN.A..T.. ...V...... .I.HDN.... ....T.S.RV.N DKRA.P..Q.
Sgc2 NRN.AT.L.. ..I....... .I.HS.T... .... I.K.Y.K..D G.RA.P..Q.
Bgc1 ERN.E..T.. ...V...... .I.HDN.... ....T.S.RA.N DKRA.P..Q.
Bgc2 ERN.E..T.. ...V...... .I.HDN.... ....T.S.RA.N DKRA.P..Q.
Bgc3 DRD..I.M.. ...M.....T .IKRENT... .....N...H ..K.Y.R..R G.RV.P..Q.
E
F
G
Con NTMKT:DTYM KFSWLTV:E: SMDKEH:CIV KHE:NK:G:D QEI:FPSIK: VV:S:::TT: 120
Mgc1 D.L..N.... ....P.R A.G...R... ...N..G.A. ...F.....K .AV.TKP..-
Mgc2 D.L..KG... ....P.R A.G...S... ...N..G.A. ...F.....K .A-------.-
Mgc4 ..I..N.R.. ....T.D ..A...S... ...N..R.V. ...L..P.GK -AF------.-
Hgc1 .....N.... ....P.K .L....R... R..N..N.V. ...I..P..T ----------
Hgc2 .....N.... ....P.E .L....R... R..N..N.I. ...I..P..T D.TTVDPKDS
Sgc1 .....K.... .L.....T.N ....Q.V.V. ..QK.IG.I. ...I.....E ..T.AVT..E
Sgc2 ...N.T.... .L.....T.N ......I.V. Q..R.IR.IN ...L....NE ..S.IVP..E
Bgc1 .....K.... ....T.N ....Q.V.V. ..KK.IG.I. ...I.....E ..T.LVP..E
Bgc2 .....K.... .L.....T.N ....Q.V.V. ..KK.IG.I. ...I.....E ..T.LVP..E
Bgc3 ..V..A.... ....SGN ......I... ...K..T.DN ...L..PVNE ..T.VVT---

```
Con  :::::::::: :::::::::: :::::::::: :::::::::: :::::::::: ::C::D:::T 180
Mgc1 ---------- ---------- ---------- ---------- ---------- --.WQ.RNDV
Mgc2 ---------- ---------- ---------- ---------- ---------- --.WQ.KNDV
Mgc4 ---------IN VNPRDSVLRH ENVNNATDLE D--------- ---------- --.MKGRKDM
Hgc1 ---------- ---------- --DVITMDPK DN--------- ---------- --.SK.AND.
Hgc2 YSKDANDVTT VDPKYNYSKD ANDVITMDPK DN--------- ---------- --WSK.AND.
Sgc1 PPTTEPPTTE PP-----NDC LTDESKVTGT GSKTACLKDE REVTGDTNST KA.QEGESSA
Sgc2 SPS------- ---------DC LNQESKVTGT GSKKACLKDE SEVTADNNST KV.LE.ESN.
Bgc1 PPTTKPPTTE PPTTEPPNDC LTDESKLTGT GSKKACLKDG S----DTNST KA.LEGESS.
Bgc2 PPTAEPP--- -------NDC LTDES-ITDT GSKKACLKDG S----DTNST KA.LEGKSS.
Bgc3 --ATKPP--- -------NDG LKDKKKQVPV V--------- -------NST KA.LK.ENN.

                                TM
Con  LQLQLTNTSA :YYTYLLLLL KS:VYFAII: :C:FRRT::C C::K:S          226
Mgc1 ....F.I... -......... ..VI.L...S FSLL...SV. GNE.K.
Mgc2 ..F.F.S... -......... ..VI.L...S FSLL...SV. GNE.K.
Mgc4 ....V.T.Y. -F....I.FF ..M.HL.FVV F.L...AAMS .DDQR.
Hgc1 .L........ -..M...... ..V......T C.LL...AF. .NGEK.
Hgc2 .L........ -..M...... ..V......T C.LLG..AF. .NGEK.
Sgc1 .R....T... -......... L.T...V..I S.V....GN. .DG.I.
Sgc2 .....M.... -......... ..V...I..T S.V....GM. .DG.NL
Bgc1 .....M.... A......F.F L.T...VV.I S.V....GVV *DG.I.
Bgc2 .....M.... A......... L.T....V.I S.V....GV. .DR.I.
Bgc3 ...H.M.... -........I T.T..LV..T S.V....GV. GIQ.S.
```

Figure 2. Alignment of TCR γ constant region gene sequences. A consensus (Con) is shown where 6 or more nucleotides are identical. A ":" in the consensus sequence indicates that no consensus was found at that position. A "-" indicates a gap which was introduced to optimize the alignment. A"." denotes identity with the consensus.

```
Con   A:A:AAACCT TGATGCAGAC ATTTCCCCCA AGCCCACTAT TTTTCTTCCT TC:ATTGCTG AAA:AAA:C: C:ATAA:GCT GGAACATACC    9 0
Mgc1  .C.A..GG.. .......... .......... .......... ...C...... ..TG...... ...C...T.T .C....GA.. ..G.......
Mgc2  .C.A..GG.. .......... .......... .......... .......... ..TG...... ...C...T.T .C....GA.. ..G.......
Mgc4  .C.A.CG.AC ...CT..... T....T.... ....T..... ...C...... ..TGC..... ...C...T.T .C....G... ..G..C....
Hgc1  .T.A.C.A.. .........T G......... .......... .......... ..A....... ...C...G.T .C.G..G... ..........
Hgc2  .T.A.C.A.. .........T G......... .......... .......... ..G....... ...C...A.T .C.G..G... ..........
Sgc1  .G.G...... ..C....... .C........ ........G. .......... ..A.....C. ...TC..C.A TG....T... ........T.
Sgc2  .T.G...... ..C.A..... C....A.... ......T... .......... ..A....... ...TC..C.A .AG...AA.. ..........
Bgc1  .G.G...... C..G...... .C........ ........G. .......... ..A....... ...TC..C.A TG....T... ........T.
Bgc2  .G.G...... C..G...... .C........ ........G. .......... ..A....... ...T...C.A TG....T... ........T.
Bgc3  .T.G.G.... ....AT.... ..G....... .......... G......... ..G...A... ...TC..G.G TG.G..TA.. ........T.

Con   TTTGTCTCCT :GAGAAATTT TTCCCTGATG TTATTA:GGT :TATTGGAAA GAAAAGAATG GCAA:A:::T TCTG::ATCC CAG:AGGGAA 1 8 0
Mgc1  .......... T..A..G..C ..T..A.... ....A.G... G......... ......G... ....T.CTA. CT..GAC... ...G.A..GG
Mgc2  .......... T..A..G..C ..T..C.... .C..A.G... G......... .......... ....T.CTA. C...GAC... ...G.A..GG
Mgc4  .......... T..A..G..C ..T...A.A. .C..A.G... G......... ......G... .TG.A.AGA. ....GA.... ...G.A....
Hgc1  .......T.. T......... .......... ......A.A. AC.....C.. ........GA ....C.CGA. ....GG.... ...G....G.
Hgc2  .......T.. T......... .....A...A ......A.A. AC.....C.. ........GA ....C.CGA. ....GG.... ...G....G.
Sgc1  .......... G......... .......... .G....CA.. G.C.....G. .T...A.... A...A.GAGC ....CC.... ...C......
Sgc2  .......T.. G......... .........A ......A... T......... ......G... ....C.GAGC ....CC.... ...C......
Bgc1  .......... G......... .......... ......CA.. T.C.....G. .C...A.... A...A.GAGC ....CC.... ...C......
Bgc2  .......... G......... .......... ......CA.. T.C.....G. .C...A.... A...A.GAGC ....CC.... ...C......
Bgc3  .......T.. G.....T... ......C... ......A... T.......G. .....A.GA. ....C.GAG. .T..CC.... ...C......

Con   ATACCATGAA GACTAA:GAC ACATACATGA AGTT:AGCTG GCTGAC:GTG :C:GAAAA:T C:ATGGATAA AGAACA:A:A TGTATTGTCA 2 7 0
Mgc1  ....GC.... ......C... .......... ....T..... ...T..A... C.T....GGG .A....GG.. ...G..C.G. ..........
Mgc2  ....GC.... ......G.G. .......... ....T..... ...T..T... C.T....GGG .A....GG.. ...G..C.GT ..........
Mgc4  .C.....A.. A.....T... .G........ .A..T..T.. ......G... A.T...G.T. .A....C... ......T.GT ..C.......
Hgc1  .C........ ......C... .......... .A..T..... .T.A..G... C.A.....G. .AC....C.. ......C.G. .....C....
Hgc2  .C........ ......C... .......... .A..T..... .T.A..G... C.A...G.G. .AC....C.. ......C.G. .....C....
Sgc1  .C........ ...C..G... .......... ..C.C..... ......C... A.C.....C. .T........ .C....TGTG ...G......
Sgc2  .......... C..C.CT... .......... ..C.C..... ......C... A.T.....C. .C........ ......T.TC ...G.C...C
Bgc1  .......... ...C..G... .......... ....C..... ......C... A.C.....C. .C........ .C....TGT. ...G......
Bgc2  .......... ...C..G... .......... ....A..... ......C... A.C.....C. .C........ .C....TGT. ...G......
Bgc3  .....G.... ....GCT... .......... ....C..... ......C... T.T.G...C. .C........ ......C.T. .....C....
```

```
Con   AACATGAGAA :AATAAA:GA GGA:TTGATC AAGAGATT:T :TTTCCTTCA ATAAA::AAG TTGTCA:C:C A:::G:::CT AC:ACT:::: 3 6 0
Mgc1  .......... C..C...G.. ...GCA.... ........T. C......... .....GA... ...CTGTGAG TACCAAGC.. ..A...----
Mgc2  .......... C..C...G.. ...GCA.... ........T. C..C...... .....GA... ...CT----- ---------- ..A...----
Mgc4  ....C..A.. C..C...A.. ...G....C. ........C. G......C.C ...GGTA..- --.CTTT.-- ---------- ..A...----
Hgc1  G......... T......AAC ...G...... ....A...A. C......C.. .....GAC.- ---------- ---------- ----------
Hgc2  G......... T......AAC ...A...... ....A...A. C......C.. .....GAC.. A.....C.A. .GTG.ATC.C .AAGACAGTT
Sgc1  ....CC.... A....T.G.. ...A...... ........A. T......... .....AG... ......C.T. .GCT.TCA.. ..T...GAGC
Sgc2  .........G A....T.A.. ...A..A... ........C. T......... .....TG... ......G.T. .ATT.TCC.. ..C...GAGT
Bgc1  ....CA.... A....T.G.. ...A...... ........A. T......... .....AG... ......C.T. .CTT.TCC.. ..T...GAGC
Bgc2  ....CA.... A....T.G.. ...A...... ........A. T......... .....AG... ......C.T. .CTT.TCC.. ..T...GAGC
Bgc3  .......... A......AC. ...GA.A... ........C. T.....GC.. G.G..TG... ......CTT. .GTT.TCA.. ----------

Con   :::::::::: :::::::::: :::::::::: :::::::::: :::::::::: ::::A:::T: ::A::GATGA :A:CA:A:T: ::::::A::G 4 5 0
Mgc1  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Mgc2  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Mgc4  ---------- ---------- ---ATTAATG TTAATCCCAG AGATAGTGTT TTGCGGCA.G AA.AT.T.A. C.ATGCTACT GATCTGGAA.
Hgc1  ---------- ---------- ---------- ---------- ---------- ---------- -----....T C.T..C.A.G GATCCC.AA.
Hgc2  ATTCAAAAGA TGCAAATGAT GTCACCACAG TGGATCCCAA ATACAATTAT TCAA.AGA.G CA.AT....T C.T..C.A.G GATCCC.AA.
Sgc1  CTCCAACTAC TGAGCCTCCA ACTACTGAGC CTCCA----- ---------- AATG.TTG.T TG.CT..... A.G..A.G.T ACAGGT.CT.
Sgc2  CTCCCAGT-- ---------- ---------- ---------- ---------- ---G.CTG.T TA.ACC.G.. A.G..A.G.T ACCGGC.CT.
Bgc1  CTCCTACTAC TAAGCCTCCT ACTACTGAGC CTCCTACTAC TGAGCCTCCA AATG.TTG.T TG.CC..... A.G..A.C.T ACAGGT.CT.
Bgc2  CTCCTACTGC TGAGCCTCCA ---------- ---------- ---------- AATG.TTG.T TG.CT..... A.G.---A.T ACAGAT.CT.
Bgc3  -----GCTAC TAAACCTCCA ---------- ---------- ---------- ---A.TGA.G GTTTGA.G.. T.AA.A.AAA CAAGTCCCT.

Con   :T::T::::: :::::::::: :::::::::: :::::::::: :::::::::: :::::TGT:: G:AAGATGAA AA::ATAC:C 5 4 0
Mgc1  ---------- ---------- ---------- ---------- ---------- ---------- -----..CTG .C.....A.. ..TG..GTG.
Mgc2  ---------- ---------- ---------- ---------- ---------- ---------- -----..CTG .C.....A.. ..TG..GTG.
Mgc4  A.-------- ---------- ---------- ---------- ---------- ---------- -----..CAT .A...GAAG. ..GG...TGT
Hgc1  ACAA.----- ---------- ---------- ---------- ---------- ---------- -----...TC AA......C. ..TG....A.
Hgc2  ACAA.----- ---------- ---------- ---------- ---------- ---------- -----..GTC AA......C. ..TG....A.
Sgc1  G.TC.AAAAC AGCTTGTCTG AAAGACGAAC GTGAAGTCAC TGGTGATACT AATTCTACAA AGGCA...CA .G...G.... .GCAG.G.C.
Sgc2  G.TC.AAAAA AGCTTGTCTG AAAGATGAAA GCGAAGTCAC TGCTGATAAT AATTCTACAA AAGTG...CT .G........ .GCA....C.
Bgc1  G.TC.AAAAA AGCTTGTCTG AAAGATGGAA GT-------- ----GATACT AATTCTACAA AGGCA...CT .G...G.... .GCAG...C.
Bgc2  G.TC.AAAAA AGCTTGCTTG AAAGATGGAA GT-------- ----GATACT AATTCTACAA AGGCA...CT .G...G.A.. .GCAG...C.
Bgc3  T.GTA----- ---------- ---------- ---------- ---------- AATTCTACAA AAGCA...CT .A........ ..CA....C.
```

```
Con   TGCAGCTGCA GCTCACGAAC ACCTCT:::G CCTATTACAC CTACCTCCTC CT:CTCCTCA AGAG::TGGT CTACTTTGCC ATCATCA:CT  6 3 0
Mgc1  .......C.. .T......T. ..........  ....C.....  ......T...  ..G.......  ....TG..A.  ......G...  .......G..
Mgc2  .....T.C..  .T......G. ..........  ....C.....  ..........  ..G.......  ....TG..A.  ......G...  .......G..
Mgc4  .......T..  .G....C.C. ....A.GCA-  --.TC.....  .......A..  ..GT..T...  ....CA....  .C....G...  T..G.TGTA.
Hgc1  .A.T......  ......A...  ......GCA-  ----------T G.........  ..G.......  ....TG....  ...T......  .......C..
Hgc2  .A.T......  ......A...  ......GCA-  ----------T G.........  ..G.......  ....TG....  ...T......  .......C..
Sgc1  ...G......  ........C.  ..........  ..........  .......T.A  ..C......C  T...CACA..  ........TT  G......TT.
Sgc2  ..........  .....T....  ..........  ..........  A.........  ..C.......  ....TG....  .......AT.  .......C..
Bgc1  ..........  .....T....  ......GCA.  ..........  .......T.A  T.C...T..C  T...CACA..  ........TT  G......TT.
Bgc2  ..........  .....T....  ......GCA.  ..........  .......T.A  ..C......C  T...CACA..  .........T  G......TT.
Bgc3  ..........  C....T....  ..........  ..........  A.........  ..C...A.T.  C...CAC...  ....C...T.  .......C..

Con   :CTGT:TGTT TAGAAGAACA G:C:TCTGCT G::ATGGGAA GA::TC:                                              6 8 7
Mgc1  T..C.C..C. ..........  TCTG....TG .CA...A...  ..AA..C
Mgc2  T..C.C..C. ..........  TCTG....TG .CA...A...  ..AG..C
Mgc4  T....C.A.. C......G..  .C.A.G.C..  .TG...ATC.  ACGT..A
Hgc1  G....C..C. .........G  .CTT......  .CA....AG.  ..AA..A
Hgc2  G....C..C. .G.......G  .CTT......  .CA....AG.  ..AA..A
Sgc1  C....G.... ...G.....C  .G.AA.....  .TG.C.....  ..TC..G
Sgc2  C....G.... .C.G......  .G.A.G....  .TG.......  ..AC.TG
Bgc1  C....G.... ..........  .G.G..GT..  .AG.C.....  ..TC..G
Bgc2  C....G.... ...G......  .G.G......  .TG.CA....  ..TC..G
Bgc3  C....G.... C..G......  .G.G.....G  .CAT.CA...  ..GC..G
```

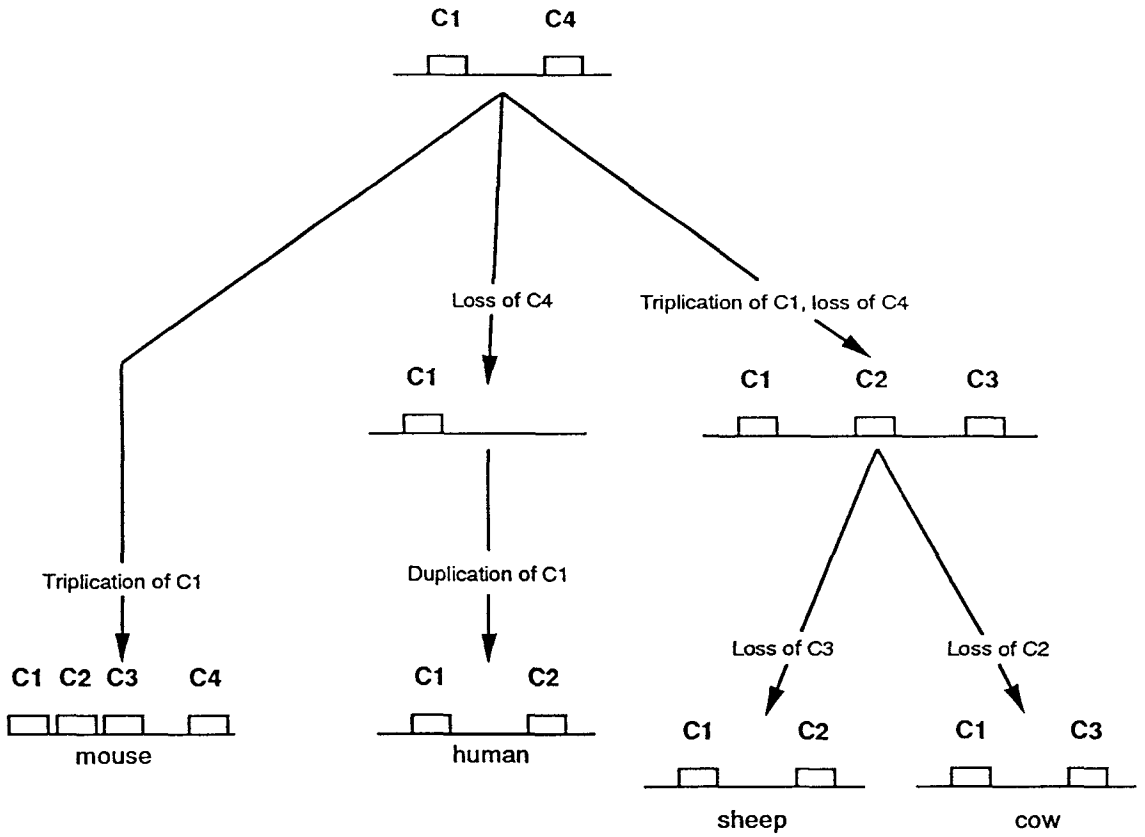


Figure 3. Proposed evolution of the Cγ genes of mouse, human, sheep and cow. The ancestral locus is shown as containing 2 Cγ genes and the bovine Cγ2 gene is assumed to be an allelle of bovine Cγ1 (and therefor not shown). See text for a full explanation.

Figure 4. Alignment of TCR Vγ region protein sequences. A consensus (Con) is shown where 9 or more amino acids (60%) are identical. A ":" in the consensus sequence indicates that no consensus was apparent. A "-" indicates a gap which was introduced to optimize the alignment. A"." denotes identity with the consensus. The putative ß strands are underlined.

```
                    A        A'            B                            C          C'
Con     ::::::::LE:  :::SV:::::  ::A:I:C:::  :::::::::I  HWY:::::::  :::L:Y::::   60
Mgv1.1  G---LGQ..Q  TEL..TRETD  ES.Q.S.IVS  LPYFS-NTA.  ...RQKAKK-  FEY.I.VSTN
Mgv1.2  G---LGQ..Q  TEL..TRETD  ENVQ.S.IVY  LPYFS-NTA.  ...RQKTNQQ  FEY.I.VATN
Mgv1.3  G---LGQ..Q  TEL..TREQD  ES.Q.S.IVS  LPYFS-NTA.  ...RQKPNQQ  FEY.I.VETN
Mgv2    G---HGK..Q  PEI..SRPRD  ET.Q.S.KVF  IESFR-SVT.  ...RQKPNQG  LEF.L.VLAT
Mgv3    G---DSWISQ  DQL.FTRRPN  KTVH.S.KLS  GVPLH-NTI.  ...QLKEGEP  LRRIF.GSVK
Mgv4    G---TSLTSP  LGSY.IKRKG  NT.FLK.QIK  TSVQKPDAY.  ...QEKPGQR  LQR.L.SSSK
Mgv5    R-QTSSN..E  RIM..TKLEG  SS.IMT.DTH  RTGT----Y.  ...RFQKGRA  PEH.L.YNFV
Hgv1.2  S-QKSSN..G  RTK..IRQTG  SS.E.T.DLA  EGSNG---Y.  ...LHQEGKA  PQR.Q.YDSY
Hgv1.3  S-QKSSN..G  RTK..TRQTG  SS.E.T.DLT  VTNTF---Y.  ...LHQEGKA  PQR.L.YDVS
Hgv1.4  S-QKSSN..G  RTK..IRQTG  SS.E.T.DLA  EGSTG---Y.  ...LHQEGKA  PQR.L.YDSY
Hgv1.5  S-QKSSN..G  RMK..TRPTG  SS.E.T.DLT  VINAV---Y.  ...LQQEGKT  PQH.LHYEVS
Hgv1.8  S-QKSSN..G  RTK..TRPTG  SS.V.T.DLP  VENAV---YT  ...LHQEGKA  PQR.L.YDSY
Hgv9    CVYGAGH..Q  PQI.STKTLS  KT.RLE.VVS  GIKIS-ATS.  Y..RERPGEV  IQF.VSI-SY
Hgv10   G-LGLSKV.Q  FQL..STEVK  KSID.P.KIS  STRFE-TDV.  ...RQKPNQA  LEH.I.IVST
Hgv11   A---LGQ..Q  PEI..SRPAN  KS.H.SWKAS  IQGFS-SKI.  ...WQKPNKG  LEY.LHVFLT

                           D           E                      F
Con     ::::::::::  :::K::::::  ::::::::::  ::::::D:::  YYCA:W::             108
Mgv1.1  YNQR--PLGG  KNK.IEASKD  FQTSTSTLKI  NYLKKE.EAT  ....V.I-
Mgv1.2  YNQR--PLGG  KHK.IEASKD  FKSSTSTLEI  NYLKKE.EAT  ....V.M-
Mgv1.3  YNQR--PLGG  KNK.IEASKD  FQTSTSTLEI  NYLKKE.EAT  ....V.I-
Mgv2    PTHI--FLDK  EYK.MEASKN  PSASTSILTI  YSLEEE.EAI  ...SY---
Mgv3    TYKQ----DK  SHSRLEIDEK  -DDGTFYLII  NNVVTS.EAT  ....C.--
Mgv4    ETIV-YEKDF  SDERYEARTW  QSDLSSVLTI  HQVTEE.TGT  ....C.--
Mgv5    SSTTVVDSRF  NSE.YHVYEG  PDKRYKFVLR  NVEES-.SAL  ....S.--
Hgv1.2  NSKVVLESGV  SPG.YYTYAS  TRNNLRLILR  NLIEN-.SGV  ....T.DG
Hgv1.3  TARDVLESGL  SPG.YYTHTP  RRWSWILRLQ  NLIEN-.SGV  ....T.DR
Hgv1.4  TSSVVLESGI  SPG.YDTYGS  TRKNLRMILR  NLIEN-.SGV  ....T.DG
Hgv1.5  NSRDVLESGL  SLG.YYTHTP  RRWSWNLRLQ  NLIEN-.SGV  ....T.GR
Hgv1.8  NSRVVLESGI  SRE.YHTYAS  TGKSLKFILE  NLIER-.SGV  ....T.DR
Hgv9    DGTVRKESGI  PSG.FEVDRI  PETSTSTLTI  HNVEKQ.IAT  ....L.EV
Hgv10   KSAARRSMGK  TSN.VEARKN  SQTLTSILTI  KSVEKE.MAV  ....A.WV
Hgv11   ISAQ-DCSGG  KTK.LEVSKN  AHTSTSTLKI  KFLEKE.EVV  .H..C---
```

Figure 5. Alignment of TCR Vγ gene sequences. A consensus (Con) is shown where 9 or more nucleotides (60%) are identical. A ":" in the consensus sequence indicates that no consensus was found at that position. A "-" indicates a gap which was introduced to optimize the alignment. A"." denotes identity with the consensus.

```
Con     :::G:::::: ::::::T::: :A::TGGA:C A:A:::::AT :TC:GTCACC AG:::GAC:G :::A:ACTGC ::AAAT::CT TG T:A::TT:    9 0
Mgv1.1  TTG.G----- -----CT.GGG C.GC....G. .A.CTGAAT. A..G...... ..AGA...A. ATG.G.G... GC....AT.C ...ATAG..T
Mgv1.2  TTG.G----- -----CT.GGG C.GC....G. .A.CTGAAT. A..G...... ..AGA...A. ATG.G.A..T GC....AT.C ...ATAG..T
Mgv1.3  TTG.G----- -----CT.GGG C.GC....G. .A.CTGAAT. A..G...... ..AGA.CAA. ATG.G.G... GC....AT.C ...ATAG..T
Mgv2    TTG.A----- -----CA.GGG A.GT....G. .ACCTGAA.. A..AA.TT.. ..ACCA.GA. ATG.G..... AC....AT.C ...A.AG..T
Mgv3    ATG.A----- -----GACTCC TGGA.ATCT. .GGATCAGC. C..CT.T... C.AAGAC.AA ACA.G..G.T GC.C..AAG. ..CA.GC.CT
Mgv4    AGG.A----- -----ACGAGT CTCAC.TCA. CTCTGGGGTC A.AT....T. .AGAG..AA. GAA.T..G.. TTTTC.CAAA ...C.AA.AA
Mgv5    GCA.A---CA AACATCCTCC A.CT....AG .A.GAATA.. G..AA..... .AGCTAGAG. GGTCCT.... TAT...GA.. ...G.TAC.C
Hgv1.2  CCA.T---CA GAAATC.TCC A.CT....AG GG.GAACG.A G..A....T. ..GCA...T. GGTCAT.... TG....CA.. ...G.TC..G
Hgv1.3  CCA.T---CA GAAATC.TCC A.CT....AG GG.GAACG.A G..A...... ..GCA...T. GGTCAT.... TG....CA.. ..CG.TC..A
Hgv1.4  CCA.T---CA GAAATC.TCC A.CT....AG GG.GAACG.A G..A....T. ..GCA...T. GGTCAT.... TG....CA.. ...G.TC..G
Hgv1.5  CCA.T---CA GAAATC.TCC A.CT....AG GG.GAATG.A G..A...... ..GCC...T. GGTCAT.... TG....CA.. ...G.CC..A
Hgv1.8  CCA.T---CA GAAATC.TCC A.CT....AG GG.GAACA.A G..A...... ..GCCA..T. GGTCAT.A.. TGT...CA.. ...G.TC..C
Hgv9    TGT.TGTATA TGGTGCAGGT C.CC.A..G. .ACCTCAA.. T..CAGT..T .AAAC.CTGT CAA.A..A.. CCGCC.GGAA ...GTGG.GT
Hgv10   TTG.A---CT TGGATTATCA A.AG....G. .GTTCCAGC. A..CA.TT.. .CGGAAGTCA AGA.A.G.AT TG.C..AC.. ..CA.GA.AT
Hgv11   TTGCA----- -----CT.GGG C.GT....A. .ACCTGAA.. A..TA.TT.. ..ACCAG.AA ATA.G.G... CC.C..AT.. ..GA.GGCAT

Con     CT:::::A:: T:::::C::: ::::::::CA T:CACTGGTA CC:ACA:AA: ::AGG::AGG :C:TA:AG:: TCT::T:TA: ::T::::C::  1 8 0
Mgv1.1  ..CTTCC.TA .TTCTC.--- AACACAGCT. .A..T..... ..GG..A..A GC.AAAA..- --T.TG..TA ...AA.A..T GTCTCAA.AA
Mgv1.2  A.CTTCC.TA .TTCTC.--- AACACAGCT. .A..T..... ..GG..A..A AC.AATC.AC AGT.TG..TA ...AA.A..T GTCGCAA.AA
Mgv1.3  ..CTTCC.TA .TTCTC.--- AACACAGCT. .A..T..... ..GG..A..A CC.AATC.AC AGT.TG..TA ...AA.A..T GTCGAGA.AA
Mgv2    TCATCGA.AG CTTTAGG--- AGTGTAAC.. .A........ ..GG..G..A CC.AACC.A. GTT..G..TT ...AT.A..T GTCCTTG.AA
Mgv3    ..GGGGTTCC CCTTCAT--- AACACCATTG .G........ ..A..TG..A GA...GG..C C.C.GAGACG AA.CT.C..T GGCTCAGTCA
Mgv4    AAACAAGTGT .CAGAAGCCC GATGCATA.. .A........ ..A.G.G..G CC...CC..C GTC.CC.AAG AA.GC.G.GT AG.TCTT.AA
Mgv5    ACAGAAC.GG CACT------ ------TA.. .C........ ..G.TTCC.G AA...GAG.. C.CC.G..CA C..TC.C..C TA.AACTTCG
Hgv1.2  ..GAAGG.AG .AACGG.--- ------TA.. .C........ ..T...CC.G GAG..GA... C.CC.C..CG ...TCAG..C TA.GACT.CT
Hgv1.3  ..GTAAC.AA .ACCTT.--- ------TA.. .C........ ..T...CC.G GAG..GA... C.CC.C..CG ...TC.G..C TA.GACGTCT
Hgv1.4  ..GAAGG.AG .ACCGG.--- ------TA.. .C........ ..T...CC.G GAG..GA... C.CC.C..CG ...TC.G..C TA.GACT.CT
Hgv1.5  ..GTAAT.AA .GCCGT.--- ------TA.. .C........ ..T...GC.G GAG..GA..A C.CC.C..CA ...TC.GC.C TA.GAAGTCT
Hgv1.8  ..GTAGA.AA .GCCGT.--- ------TA.. CC........ ..T...CC.G GAG..GA... C.CC.C..CG ...TC.G..C TA.GACT.CT
Hgv9    ..GGAAT.AA AATTTCT--- GCAACATCTG .AT.T..... T.G.G.G.GA CCT..TG.A. T.A..C..TT C..GG.G.CC AT.---T.AT
Hgv10   .GAGCAC.AG GTTTGAA--- ACAGATGT.. .T........ ..GG..G..A CC.AATC... CTT.GG..CA C..GA.C..T AT.GTCT.AA
Hgv11   .CATCCA.GG CTTTAG.--- AGTAAAAT.. .A........ .TGG..G..A CC.AACA.A. G.T..G.ATA .T.AT.AC.T GTCTTCTTGA
```

```
Con     ACAAC:C:A: :::::::::: :::T:AGG:: ::A::::::A :AAAT:TGA: GC:::TA:A: ::::::::A: :::::C:T:A :T: :T::::A 2 7 0
Mgv1.1  ..T..AATCA ACGA------ CCC.T...AG GG.AGAACA. A...AT...A ..AAG..A.G ATTTTCAA.C TTCTA.C.C. AC C T.GAAA.
Mgv1.2  ..T..AATCA ACGA------ CCC.T...AG GG.AGCACA. A...AT...A ..AAG..A.G ATTTTAAA.G TTCTA.C.C. AC C T.GGAA.
Mgv1.3  ..T..AATCA ACGA------ CCC.T...AG GG.AGAACA. G...AT...A ..AAG..A.G ATTTTCAA.C TTCTA.C.C. AC C T.GGAA.
Mgv2    C.CCTA.CCA TATT------ TTC.T..ATA AGGAGTACA. G...ATG..G ..AAG..A.A ATCCTAGTGC TTCTA.A.CG A. A T.GACA.
Mgv3    .A.CTTAC.A ACAA------ -------.ACA AGTCCCACTC CCGC.TG..A ATTGA.GAGA AG---GATGA TGGTA.C.TT TA C C.GATA.
Mgv4    .AG.AA.A.T TGTC---TAT GAGAA..ATT TT.GTGACG. A.G..A...G ..AAGG.C.T GGCAGAGTGA TTTGT.T.C. G. C C.CACC.
Mgv5    T..GTT.C.C AACTGTGGTG GAT.CCA.AT TC.ATTCAG. G....A.C.T .TTTA.GA.G GCCCGGAC.A GAGGTATAA. T. T G.GCTTC
Hgv1.2  .....T.C.A GGTTGTGTTG GAA.C...AG TC.GTCCAGG G..G.A.T.T A.TTACGC.A GCACAAGG.A CAACTTGAG. T. G A.ACTGC
Hgv1.3  C..C.G.A.G GGATGTGTTG GAA.C...AC TC.GTCCAGG A..G.A.T.T A.TCA..C.C CCAGGAGGTG GAGCTGGAT. T. G AGACTGC
Hgv1.4  ...C.T.C.G CGTTGTGTTG GAA.C...AA TC.GCCCAGG G..G.A...T A.TTA.GG.A GCACAAGG.A GAACTTGAG. A. G A.ACTGC
Hgv1.5  C....T.A.G GGATGTGTTG GAA.C...TC TC.GTCTTGG A..G.A.T.T A.TCA..C.C CGAGGAGGTG GAGCTGGAAT T. G AGACTGC
Hgv1.8  .....T.C.G GGTTGTGTTG GAA.C...AA TC.GTCGAG. A..G.A.C.T A.TTA.GC.A GCACAGGG.A GAGCCTTAA. T. T A.ACTGG
Hgv9    .TG..GGC.C TGTCAGAAAG GAA.CT..CA TTCCGTCAGG C....T...G .TGGA..GGA TACCTGAA.C GTCTA.A.CC AC T C.CACC.
Hgv10   CA..AT.CGC AGCTCGACGC AGCATG..TA AG.CAAGCA. C...GTG..G ..AAGA.AGA ATTCTCAA.C TCTCA.T.C. A. C C.TACC.
Hgv11   CA.T.T.TGC TCAA---GAT TGC.C...TG GG.AGACTA. G...CT...G .TAAG..A.A ATGCTCAC.C TTCCA.T.CC AC T T.GAAA.

Con     TAAAT::::T :GA:AAA::: GA::::G::: TCTACTACTG TGCC::CTGG :::::::: 3 2 8
Mgv1.1  .....TACT. GA.G...GAA ..TGAA.CTA C......... ...AGT.... ATAAA---
Mgv1.2  .....TACT. GA.G...GAA ..TGAA.CCA C......... ...AGT.... ATGAG---
Mgv1.3  .....TACT. GA.G...GAA ..TGAA.CCA C......... ...AGT.... ATAAA---
Mgv2    ..T..TCCT. G..GG..GAA ..CGAA.CTA .......... .T..TA.G.- --------
Mgv3    .C..CAATG. T.TC.C.TCG ..TGAA.CCA CG........ ....TG.... --------
Mgv4    ..C.CCAAG. GACAG..GAG ..CACG.GAA CT..T..... ...ATG.... --------
Mgv5    GG...GTGGA G..GTCC--- ..TTCT.CTC .G........ ....TC.... --------
Hgv1.2  G....CTAA. T..A..T--- ..CTCT.GGG ....T..... ....AC.... GACGGG--
Hgv1.3  A....CTAA. T..A..T--- ..TTCT.GGG ....T..... ....AC.... GACAGG--
Hgv1.4  G....CTTA. T..A..T--- ..CTCT.GAG ....T..... ....AC.... GATGGG--
Hgv1.5  A....CTAA. T..A..T--- ..TTCT.GGG ....T..... ....AC.... GGCAGG--
Hgv1.8  A....CTAA. T..ACGT--- ..CTCT.GGG ....T..... ....AC.... GATAGG--
Hgv9    .TC.CAATG. A..G...CAG ..CATA.CTA C......... ....TTG... GAGGTG--
Hgv10   .C..GTCCG. A..G...GAA ..CATG.CCG .T........ ...TGCG... TGGGTGGC
Hgv11   ....GTTCT. A..G...GAA ..TGAG.TGG .G...C.... ....TGT--- --------
```

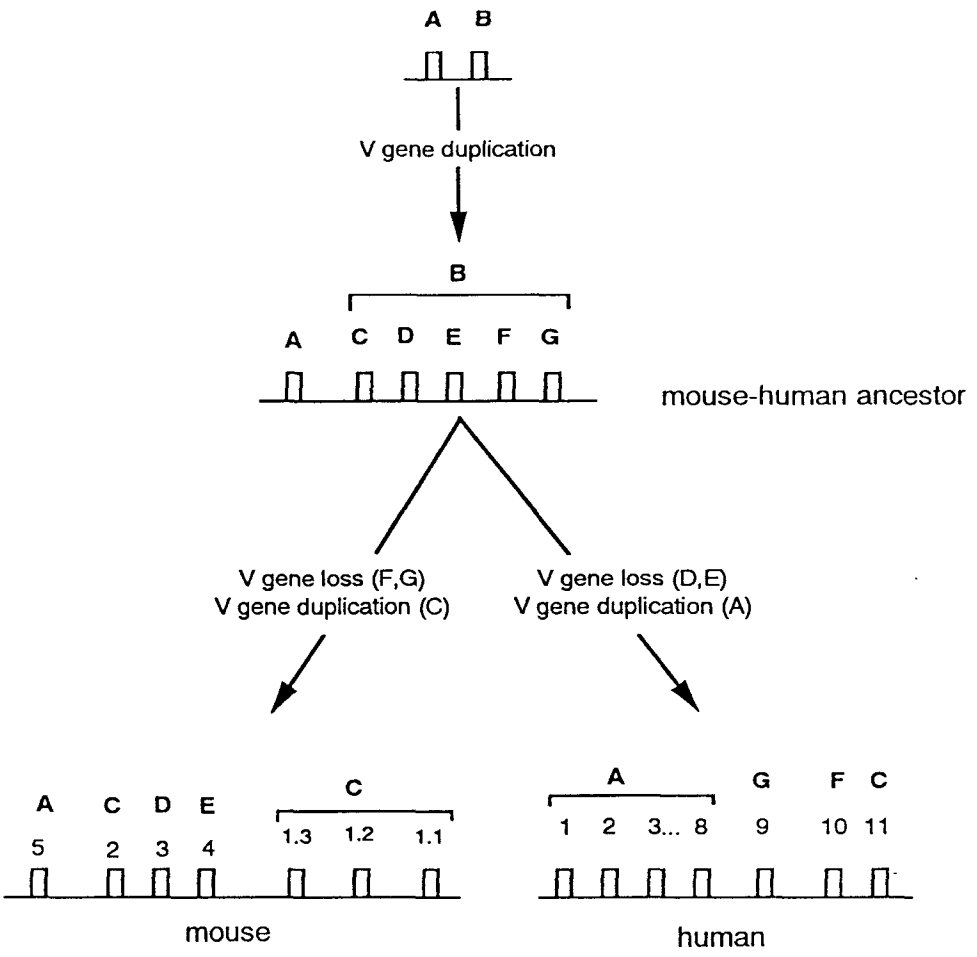


Figure 6. Proposed evolution of the functional Vγ genes of mouse and human. The ancestral locus is shown as containing 2 Vγ genes, VA and VB. The VB gene is postulated to have duplicated, resulting in a mouse-human ancestor which 5 related Vγ genes (C-G). Not all members of the human Vγ1 family are shown. The bold characters indicate V genes and the numerical characters show the current Vγ gene nomenclature for mouse and human.

Table 2. Percent identity between Cγ proteins. The values below the diagonal show the identity between the entire proteins. The values above the diagonal show the percent identity between the proteins outside the hinge region (amino acids 1-110 and 181-226). Symbols as in figure 2.

minus hinge

total amino acid

		mouse			human		sheep		bovine		
		C1	C2	C4	C1	C2	C1	C2	C1	C2	C3
mouse	C1		96	70	76	74	62	67	60	63	63
	C2	93		68	72	71	63	63	60	63	62
	C4	59	58		65	64	57	61	55	56	57
human	C1	70	66	53		97	65	65	62	65	64
	C2	68	64	51	82		65	67	62	64	62
sheep	C1	46	46	40	46	46		77	91	93	68
	C2	44	41	45	50	48	71		73	77	72
bovine	C1	42	44	41	45	46	85	68		96	71
	C2	49	49	41	49	47	86	73	89		69
	C3	52	52	44	54	43	56	60	59	59	

Table 3. Percent identity between Cγ gene sequences. The values below the diagonal show the identity between the entire genes. The values above the diagonal ("minus hinge") show the percent identity between the genes outside the hinge region (nucleotides 1-329 and 540-687).

minus hinge

	nucleo tide		mouse			human		sheep		bovine		
			C1	C2	C4	C1	C2	C1	C2	C1	C2	C3
total	mouse	C1		97	78	81	80	72	75	73	74	74
		C2	95		73	80	79	73	75	73	74	73
		C4	66	67		74	73	70	72	70	70	69
	human	C1	77	75	62		98	74	78	74	75	77
		C2	77	74	61	83		73	78	74	77	76
	sheep	C1	55	53	54	56	58		85	95	95	80
		C2	61	60	58	64	61	81		85	85	85
	bovine	C1	55	55	56	57	60	91	78		98	82
		C2	60	59	56	58	60	90	81	93		82
		C3	63	62	57	66	63	70	74	70	73	

Table 4. A. Similarity between human Vγ region segments. B. Similarity between mouse Vγ region segments.

Values above the diagonal are percent amino acid identity.

Values below the diagonal are percent nucleotide identity.

A.

HUMAN	HUMAN V1.2	HUMAN V1.3	HUMAN V1.4	HUMAN V1.5	HUMAN V1.8	HUMAN V9	HUMAN V10	HUMAN V11 (amino acid)
V1.2		75	91	73	77	23	24	22
V1.3	88		76	85	73	22	23	23
V1.4	95	88		66	78	24	24	23
V1.5	85	94	84		71	23	23	27
V1.8	89	87	89	85		25	23	24
V9	38	37	40	36	39		34	32
V10	39	37	39	37	38	48		43
V11	33	34	35	35	35	45	59	

nucleotide

B.

MOUSE	MOUSE V1.1	MOUSE V1.2	MOUSE V1.3	MOUSE V2	MOUSE V3	MOUSE V4	MOUSE V5 (amino acid)
V1.1		86	92	48	32	22	23
V1.2	94		90	49	32	21	21
V1.3	95	95		51	31	23	23
V2	68	65	72		27	28	22
V3	45	45	44	44		31	18
V4	45	41	40	44	29		20
V5	35	34	35	39	35	36	

nucleotide

Table 5. A. Similarity between mouse and human Vγ protein sequence. B. Similarity between mouse and human Vγ gene segments. The values are expressed as percent identity.

A.

HUMAN \ MOUSE	V1.1	V1.2	V1.3	V2	V3	V4	V5
V1.2	25	22	23	19	21	26	44
V1.3	26	24	25	20	23	25	45
V1.4	25	22	23	20	21	26	47
V1.5	26	24	26	21	20	23	43
V1.8	26	23	24	21	21	25	49
V9	33	30	34	31	29	26	25
V10	41	39	43	44	29	27	27
V11	51	46	51	51	28	19	22

Amino acid

B.

HUMAN \ MOUSE	V1.1	V1.2	V1.3	V2	V3	V4	V5
V1.2	36	35	35	36	33	41	65
V1.3	37	37	35	35	33	39	64
V1.4	36	36	35	38	34	40	66
V1.5	36	36	35	35	34	39	62
V1.8	36	35	35	36	33	39	68
V9	45	45	46	46	40	43	38
V10	51	51	53	56	44	41	40
V11	65	63	65	71	44	38	41

Nucleotide

Appendix

THE MAKING OF A MAP.

INTRODUCTION.

This appendix describes the progress in the mapping of the murine TCR γ locus. It emphasizes several aspects of the project: the strategies used at the various stages of the project, problems or stumble blocks that were encountered and the strategies used to overcome these. Finally, some general aspects of mapping are discussed.

STRATEGIES.

Several complementary strategies were used for mapping and cloning of the mouse TCR γ locus. The size of the locus was estimated by field inversion gel electrophoresis (FIGE). Cosmids were isolated using several γ genes and gene segments as probes. The first set of cosmids fell in 4 overlapping sets (so called contigs), with 3 gaps between the sets. The presence of rare cutting restriction enzymes in the cosmids was determined. Based on these data and the FIGE data, the sizes of the gaps were estimated.

The next attempt was to isolate (a) YAC clone(s) which contained the locus, or part of it. This did not allow us to close all gaps. Next, we screened a larger cosmid library with more probes. One cosmid was isolated which linked 2 contigs.

The other 2 contigs appeared not to be linked. As a final step, cosmid ends were sequenced in an attempt to isolate

single copy probes from the ends of contigs. These could be used to isolate cosmids which would extend further into the gaps, or even link the contigs together.

RESULTS AND DISCUSSION.

1. The size of the mouse TCR γ locus.

BALB/c liver DNA was digested with infrequent cutting enzymes, run on a field inversion gel and blotted onto Zetapobe (BioRad). The blot was hybridized with various mouse γ probes. The Cγ4 gene has limited homology to the other 3 Cγ genes and will not show up when the blot is hybridized with a Cγ1, 2 or 3 probe and washed with 0.2 x SSC. This allowed discrimination between the Cγ4 and Cγ1, 2 and 3 genes.

The most informative enzyme was Ksp I, which allowed the identification of 2 large DNA fragments, 360 and 100 kb in size, which contained all known TCR γ genes and gene segments. These 2 fragments were shown to be contiguous in subsequent cosmid cloning experiments (see below).

A problem that became apparent in the final stage of the project was that several FIGE fragments were not complete, but partial digestion products. Comparison with the FIGE map of Woolf *et al.* (1988) indicated that discrepancies existed between the sizes of the Sal I fragments in these 2 maps. As both studies used BALB/c DNA, this could not be due to strain polymorphisms. Since the sizes of the FIGE fragments in both

studies were not accurate (data not shown and see chapter III), it was not realized that some FIGE fragments in our study were partials.

2. Cosmid isolation.

The postulated map of the mouse γ locus (figure 4 in chapter I) showed that the Cγ genes and Vγ1 gene segments were interspersed in the locus. Probes for these genes were used to screen 30 cosmid library filters in duplicate (150,000-300,000 clones). Positive clones were picked and colony purified. Cosmid DNA was isolated, and aliquots were digested with Eco RI and Hind III, run on an agarose gel and blotted onto Zetaprobe. These blots were hybridized with various γ probes. Since the sizes of the Eco RI fragments which contain the gene segments were mostly known (from the literature), it was simple to place the cosmids on the map, as shown in figure 1.

The cosmids were also digested with the same rare cutting enzymes that were used for the FIGE experiments. Following gel elctrophoresis and blotting, the blots were hybridized with various γ probes and with oligonucleotides specific for the ends of the cosmid vector. The data obtained from these experiments were used to place the infrequent cutting enzymes on the map, as shown in figure 2. The presence of multiple infrequent cutting enzyme sites next to the cloning site in

the vector (pWE15A) was very helpful in mapping the infrequent cutting enzyme sites in the cloned DNA.

Also shown in figure 2 are the FIGE fragments that were placed on the map, based on the location of the infrequent cutting enzymes. The start and endpoints of certain FIGE fragments which spanned gaps between contigs were arbitrarily assigned. For many of these no probe was available for one end of the proposed fragment location.

The size of the FIGE fragments shown in figure 2 were used to calculate the size of the gaps between the contigs. The gap between the Cγ1 and Cγ3 contigs was estimated at 5-10kb (based on the 60 kb Sal I fragment). The gap between the Cγ2 and Cγ4 contigs was estimated to be 20 kb (based on the 60 kb Sma I fragment) and the size of the gap between the Cγ3 and Cγ2 clusters was estimated to be 90 kb (based on the Cγ2 and Cγ4 gap and the 180 kb Sal I fragment).

An initial attempt was made to isolate single copy probes from the ends of cosmids γ6 and γ14. These ends were cloned and partially sequenced. Cosmid γ6 contained a highly repeated sequence as determined by hybridization. Sequence analysis showed that it was part of an L1 repeat. Cosmid γ14 was determined to be a recombinant cosmid. The end fragment of this cosmid hybridized to a 450 kb Ksp I fragment, not to the expected 360 kb fragment.

The number of cosmids isolated from the library was much lower than expected based on cosmid isolation for other TCR loci (Kai Wang, personal communication). Hence, the cosmid library was rescreend with the same probes as before. Several new cosmid clones were isolated, as shown in figure 3. None of these cosmids linked contigs together.

3. YAC isolation.

In order to close the gaps, 2 approaches were taken. The first was to isolated (a) YAC clone(s) which would contain the γ locus. The second approach was to screen a larger cosmid library with more γ probes (see below).

A YAC library from C57BL/6 mouse DNA (S. Tilghman, Princeton) was screened with oligonucleotide primers specific for the Vγ1 gene segments and primers for the Cγ1, 2 and 3 genes. A single pool of 94 clones was positive. A colony filter lift containing DNA from these clones was screened with various γ probes. Initially, a false positive was isolated (background on the membrane). A genuine positive was isolated later. A problem with screening the YAC colony filter was that the amount of DNA on the filter was highly variable. Hybridization with a YAC vector probe was necessary in order to show that the genuine positive on the filter was very weak due to the low amount of DNA that was present (YAC pools are now screened by PCR amplification until single positive

clones are found).

This YAC clone contained only the 5' end of the mouse TCR γ locus, as indicated in figure 4. The Cγ1 gene was present in this YAC clone, but not any of the Vγ1 gene segments. By FIGE analysis, the YAC clone was shown to extend approximately 40 kb 3' to the Cla I site in the Vγ2 gene segment (data not shown). The size of the YAC clone was 500-600 kb, as determined by FIGE.

4. Rescreening of the cosmid library.

The previous cosmid filter hybridizations yielded fewer positives than expected. In addition, the positive clones showed up as weak spots on the autoradiograms, sometimes as single positives. It was concluded that the filters were in poor condition (contained small amounts of accessible DNA). Hence new duplicate filters were made from 100 cosmid library plates (500,000-1,000,000 colonies), which were screened with probes for the Cγ genes, for the Vγ1 gene segments and for the enhancer like fragments (located 3' to Cγ1, 2 and 3; see chapter II).

This screen yielded many positives. Cosmid 54.3 spanned the gap between the Cγ1 and Cγ3 clusters (see figure 5). It hybridized to a Cγ1 probe, as well as an oligonucleotide probe derived from the end of cosmid γ6. In order to verify that the linkage of the 2 contigs was correct, the region

between cos γ72 and cosmid γ6 was amplified from genomic DNA and from cosmid γ54.3, using 2 primers derived from the end of cosmid γ6 and 2 primers derived from the end of cosmid γ7 (which was also sequenced, data not shown). The amplified fragments from the cosmid DNA and the genomic DNA were identical in size and predicted a gap between cosmids γ6 and γ72.2 of 100-200 bp. The 'gap' was sequenced from the amplified fragment of cosmid γ54.3 and shown to be part of a L1 repeat. The sequence of the end of cosmid γ6, the 'gap' and the end of cosmid γ72.2 was colinear with the sequence of prototype L1 sequences in the GenBank database (data not shown). This established that the gap was 150 bp in size and that cosmid γ54.3 did not contain a deletion in this region. No other clones were isolated which spanned the gap between the Cγ1 and Cγ3 clusters.

Several cosmids were isolated which hybridized to the Cγ, the γ enhancer and/or Vγ1 probes. The cosmids which fell in the Cγ4 cluster all mapped to the same region of the cluster. Cosmid γ58 appeared to extend nearly 5 kb beyond the 5' end of cosmid γ74. However, it contained a Not I site which was not present in any of the Cγ4 containing cosmids. Hence this was a recombinant cosmid. Only cosmid γ63 extended beyond the 5' end of cosmid γ74, by approximately 1.2 kb.
One cosmid (γ73A) was mapped to the 3' end of the Cγ3 contig (see figure 5).

Four identical cosmids (represented by cosmid γ37-2 in figure 5) were mapped to the 5' end of the Cγ2 contig. Cosmid γ10.2 mapped to the 3' end of this contig (figure 5).

Recombinant cosmids.

Cosmid γ97 contains a Sal I site near the 3' end, which is not present in cosmid γ10.2. This identified one of these cosmids as a recombinant cosmid. The Sal I and Sma I sites in the 5' half of cosmids γ2 and the γ37-2 like cosmids did not coincide, indicating that here too a recombinant cosmid could be present. Cosmid γ37-2 was unstable, which made the interpretation of the digests more complicated.

A cosmid isolated in this screen but characterized later (cosmid γ91, see figure 6) mapped to the 3' end of the Cγ3 contig. It did not contain the Cla I site which was present in the 3' end of cosmid γ73A, indicating that one of these cosmids was recombinant.

5. Cosmid end sequencing.

The result of the work to this point still left 2 gaps, of approximately 15 and 70 kb. In order to close these gaps, single copy probes were needed from near the ends of the contigs. For this, ends of relevant cosmids were sequenced. The obtained sequences were compared to the Genbank database and classified as repeat elements, or as unknown sequences.

Two types of repeated elements were found. Many cosmid ends contained parts of L1 repeats (indicated as **L1** in figure 6). The mouse genome contains 50,000 to 100,000 copies of this repeat in the genome. Two cosmid ends contained retroviral-like sequences, which are indicated as **R** in figure 6. Only 50-100 copies of this repeat exist in the mouse genome. The organization of the L1 and retroviral sequences in cosmid γ63 indicated that the 5' end of the retroviral sequence was joined to the 3' end of an L1 sequence.

Oligonucleotide primers were designed from the 'unknown' cosmid end sequences and used for PCR amplification of these ends (using an appropiate vector primer as a second primer). The PCR fragments were used for hybridization to genomic southern blots and cosmid blots.

The end of cosmid γ37-2 hybridized to a single Eco RI fragment on a genomic blot, but only when the blot was washed at high stringency. Surprisingly, it hybridized to the end of cosmid γ91 (see figure 7), a cosmid which was isolated in the same library screen as cosmid γ73A. This cosmid had been hard to purify and was therefore not mapped when figure 6 was made. Two primers designed from the endsequence of cosmid γ37-2 amplified the same size fragment from cosmids γ37-2 and γ91, indicating that these 2 cosmids overlapped.

The end probes of cosmid γ10.2 and γ93 hybridized to 5 identical Eco RI bands in the cosmid blots and in genomic blots when the blots were washed at high stringency. This

indicated that these probes were specific for the γ locus when the washes were performed under stringent conditions. As mentioned above, either cosmid γ73A or cosmid γ91 was recombinant. As cosmid γ91 overlapped with cosmid γ37-2, cosmid γ73A had to be recombinant. Hence, its end sequence could not be specific for the γ locus. This was confirmed when it became clear that cosmid γ10.2 was also recombinant (see below). The overlap of cosmids γ91 and γ37-2 was in agreement with the FIGE data, again indicating that cosmid γ91 was a correct (i.e., nonrecombinant) cosmid, whereas cosmid γ73A was recombinant (see below and chapter II for a discussion of the FIGE data and the final map).

The presence of 2 retroviral sequences (of which only 50 to 100 copies exist in the mouse genome) in the γ locus seemed statistically unlikely (but ofcourse, the locus appears to have been generated by duplications). Therefore, the ends of cosmids γ3, γ63 and γ74 were analyzed in more detail. The junction of the retroviral and L1 elements was amplified from cosmid γ63 and sequenced. Amplification primers were made from the obtained L1 and retroviral sequences and used for amplification of cosmid DNA. The amplified fragments from cosmid γ3 and γ63 template DNA were identical. Hybridizations with these PCR fragments showed that the organizations of the ends of both cosmids were identical. The Sma I - Cla I fragment in both cosmids were also of identical size. This

indicated that these cosmids overlapped.

However, cosmids γ97 and γ10.2 did not hybridize to these PCR probes. It was already argued above that (1) one of these cosmids was recombinant, and (2) the end of cosmid γ10.2 was not specific for the γ locus, as its crosshybridizing counterpart (the end of cosmid γ73A) was not specific for the γ locus. Furthermore, the 60 kb Sma I FIGE fragment predicted a gap of 15-20 kb between the Cγ2 and Cγ4 contigs. This did not agree with an overlap of cosmids γ3 and γ63.

However, if the Sma I fragment was a partial, then the overlap is correct. As shown in figure 7, there is a 50 kb Sma I fragment in the Cγ2-Cγ4 region, assuming that the internal Sma I site is not digested. The overlap, as shown in figure 7, also agrees with the map of Woolf *et al.* (1988), which predicted a 90 kb Sal I fragment containing both Cγ4 and Cγ2.

The Sal I fragment of 180 kb shown in figure 6 agrees with the idea that this also is a partial digestion product, and also agrees with the map of Woolf *et al.* (1988), except that they missed a 25 kb fragment (see chapter II). This correlates with an overlap of cosmids γ91 and γ37-2 and explains the discrepancies between the FIGE data of our study and the map of Woolf *et al.* (1988).

Conclusions.

Several problems were evident in this large scale mapping project.

First, the available probes were clustered in the locus and there were large regions for which no probes were available. For instance, the distance between the enhancer-like elements 3' to Cγ2 and Cγ3 is 75 kb, with no other known γ gene segments located inbetween. Hence it can be considered luck that we obtained cosmids γ91 and the 4 γ37-2 like cosmids, and thus linked the Cγ2 and Cγ3 contigs.

Second, the FIGE data were in part derived from partial digestion products. While partial digestion products are useful in large scale mapping, it is important to know what bands on an autoradiogram are partial products. The partial products led us to overestimate the sizes of the gaps between the contigs. Also, the lack of single copy probes led us to misplace the ends of the FIGE fragments on the map.

Third, several recombinant cosmids were encountered. These can only be identified when at least 2 cosmids are present which map to equivalent positions, but which are nevertheless different at one end (e.g., cosmids γ91 and γ73A). Which cosmid is correct can only be determined by other methods (e.g., the overlap between cosmids γ91 and γ37-2), or by isolating more cosmids in the relevant area. As shown in the Cγ2 - Cγ4 region, recombinant cosmids can map preferentially to certain region (cosmids γ14, γ97 and γ10.2 were all

recombinant). The absence of many cosmids in one particular region of a locus (e.g., the Cγ2 region shown in figure 7) may indicate that that region is hard to clone into cosmids. This is presumably because the partial digestion conditions which give a 'good' library cause more, or fewer than average Sau 3AI sites in this region to be digested (i.e., these sites are preferential, or poor digestion sites, respectively). Preferential digestion leads to DNA fragments which are smaller than 35-40 kb and which can not be efficiently cloned into cosmids. These smaller fragments can however ligate together and into cosmid vectors. If these products can be packaged efficiently, they can show up as recombinant cosmids.

Identification of recombinant cosmids.

The γ cosmids were mapped by digestion of cosmid DNA with infrequent cutting enzymes, and hybridization with γ probes. This also proved to be an efficient method for identifying the presence of recombinant cosmids.

Cosmid Vector.

The use of cosmid vector pWE15A, which has multiple infrequent cutting enzyme sites near the cloning site, was very useful for the cosmid mapping.

Cosmid end sequencing.

Sequencing of the ends of cosmids works fairly well with the USB cycle sequencing kit. A prerequisite appears to be decent quality cosmid DNA (data not shown). Many cosmids can be processed at the same time. A disadvantage is that the primers have to be labelled with γ^{32}P-ATP. These labelled primers and the sequencing reactions are very hot and not stable for very long. A fluorescent sequencing approach or other alternative would be desirable.

This method works well enough that it should be considered a viable method for cosmid walking. Unfortunately, a portion of mouse cosmid ends are not expected to contain single copy sequences, especially not in loci which contain repeated sequences (e.g., all TCR loci). This limits the usefulness of this technique in species wich have genomes with large numbers of repeats.

Reference.

Woolf, T., Lai, E. Kronenberg, M., and Hood, L. (1988). Mapping genomic organization by field inversion and two-dimensional gel electrophoresis: Application to the murine T-cell receptor gamma gene family. Nucleic Acids Res. *16*, 3863-3875.

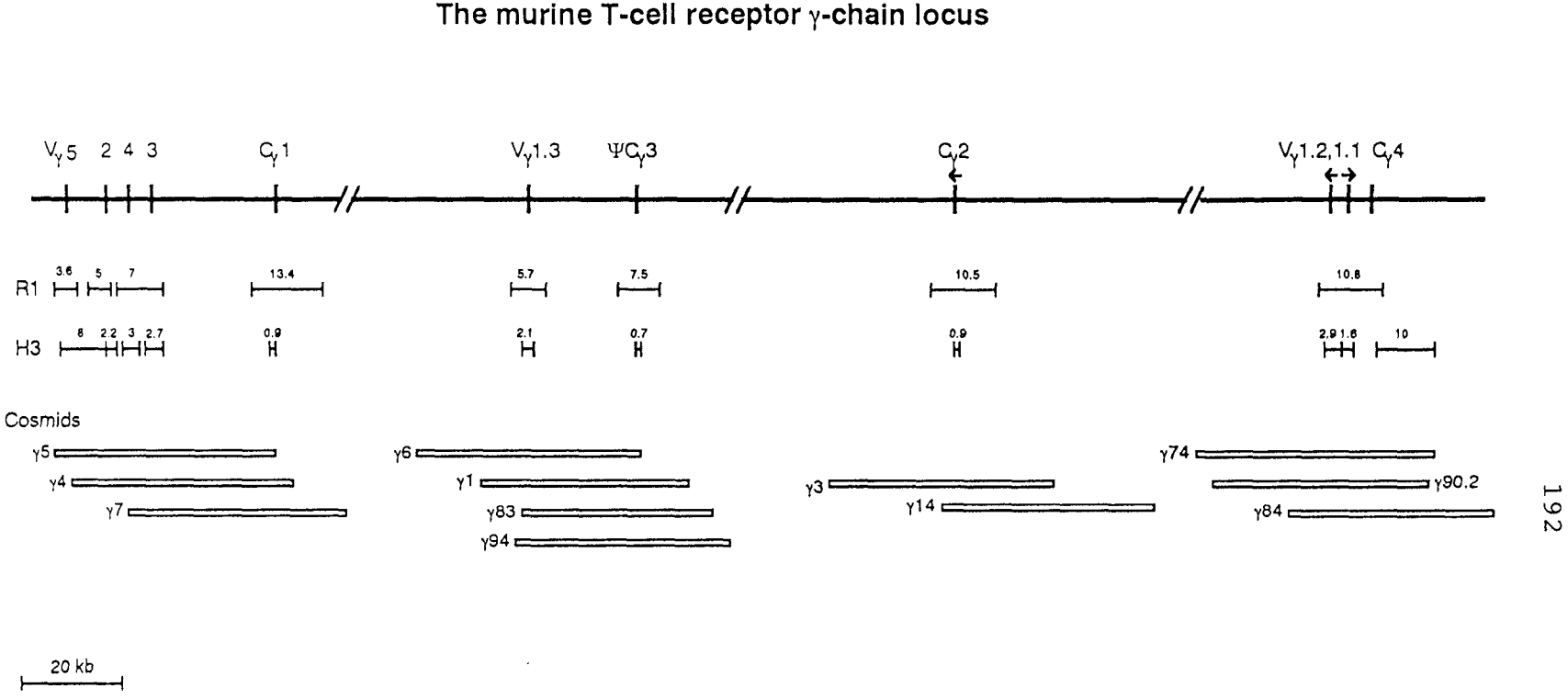


Figure 1. Preliminary map of the mouse TCR γ locus. The genes and gene segments are all indicated with vertical bars. RI and H3 indicate the Eco RI and Hind III restriction fragments that contain the genes and gene segments listed above them.

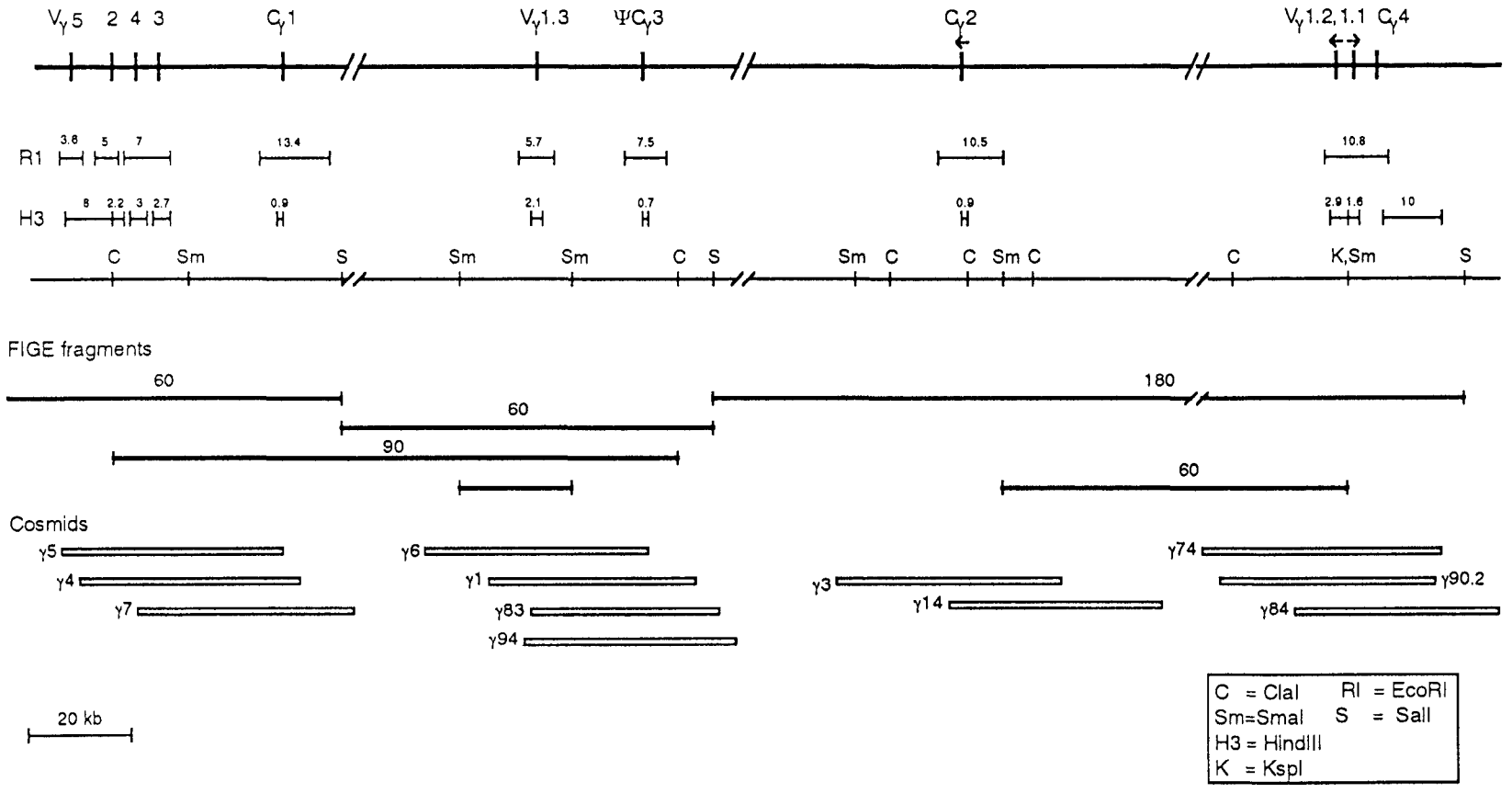


Figure 2. Preliminary map of the mouse TCR γ locus. The FIGE fragments and location of the cosmids are indicated. Infrequent restriction sites are indicated.

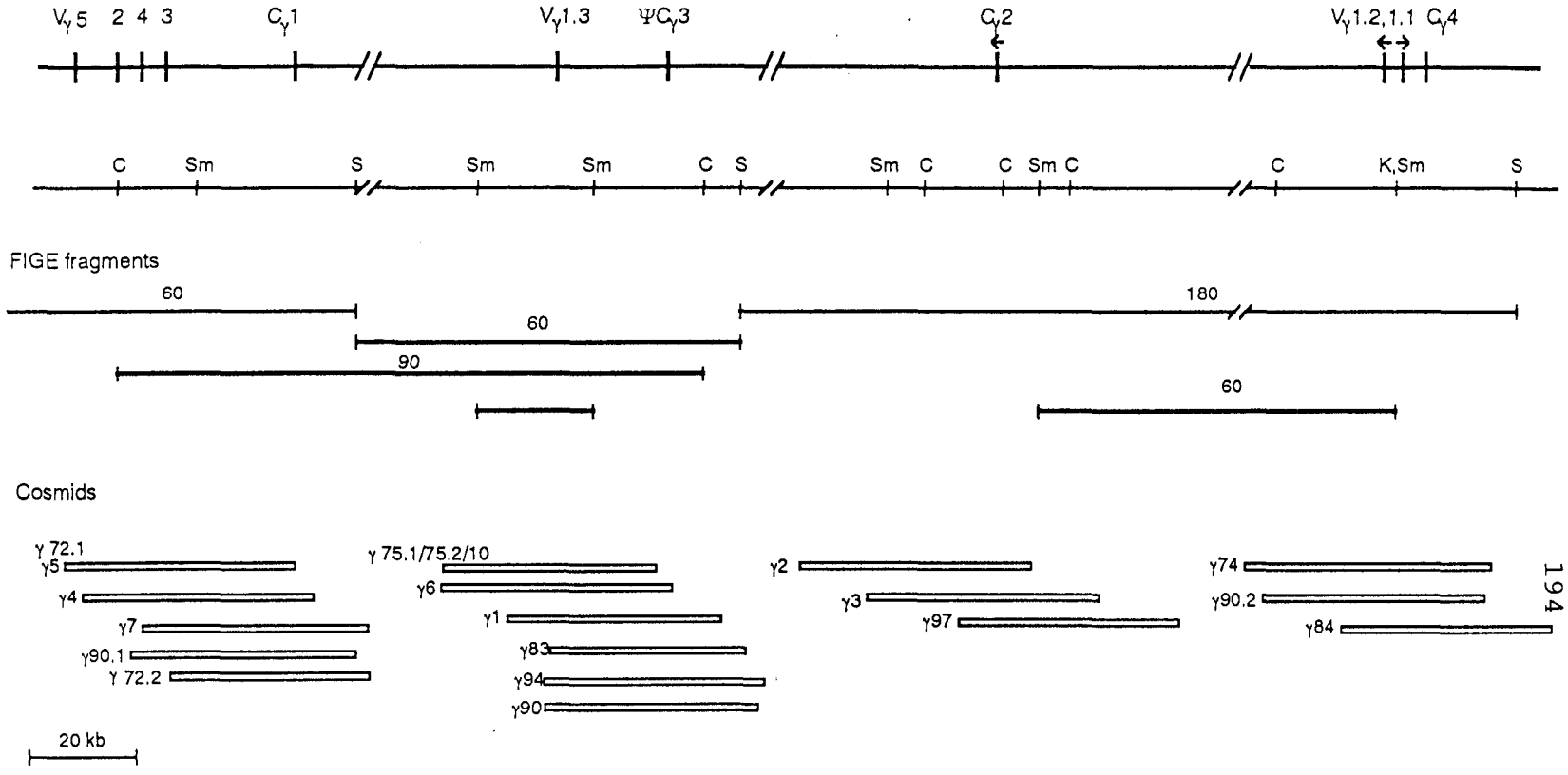


Figure 3. Preliminary map of the mouse TCR γ locus. The FIGE fragments and location of the cosmids are indicated. Symbols as described in figures 1 and 2.

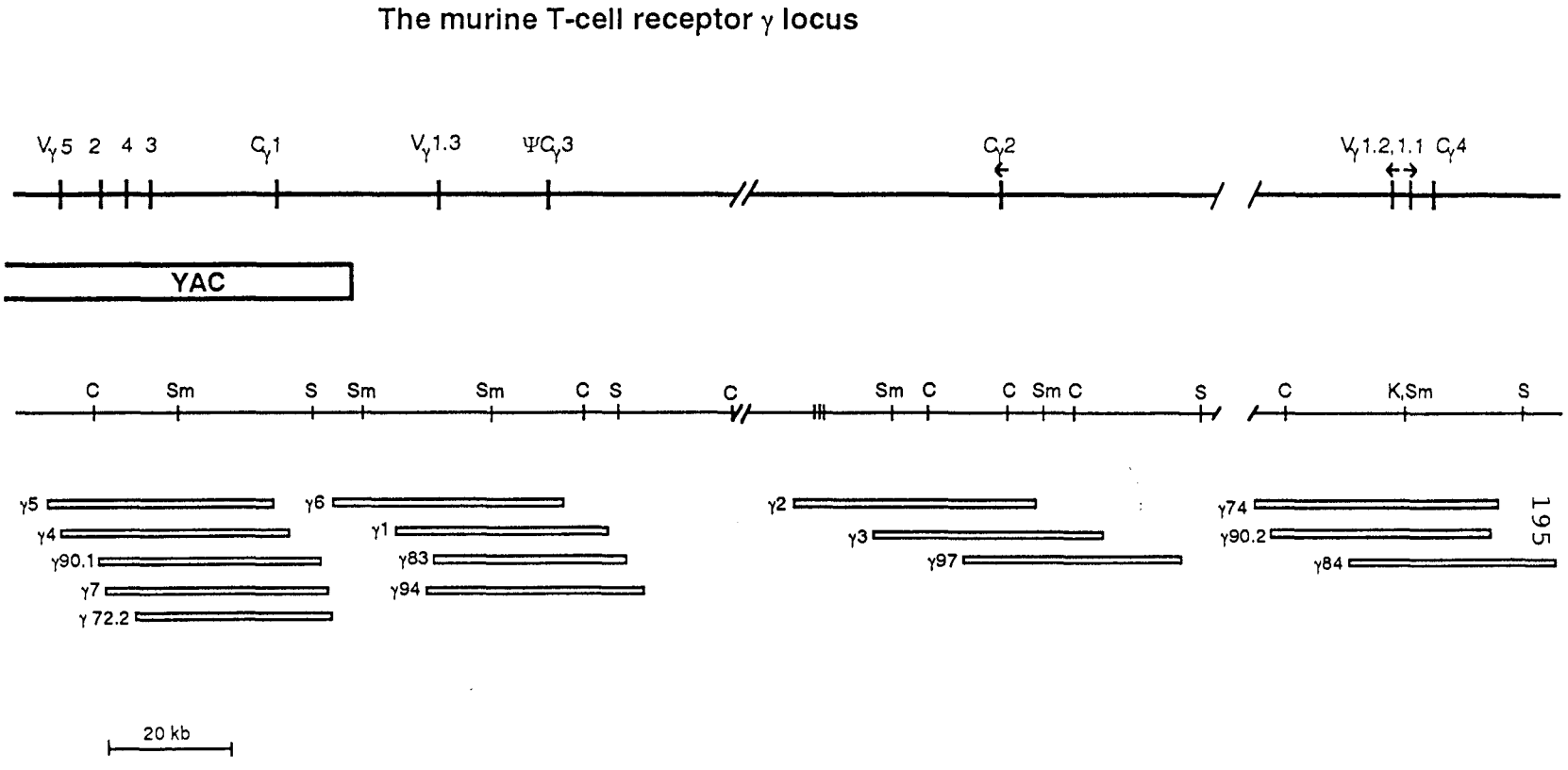


Figure 4. Preliminary map of the mouse TCR γ locus. The approximate location of the YAC clone is indicated. Symbols as described in figures 1 and 2.

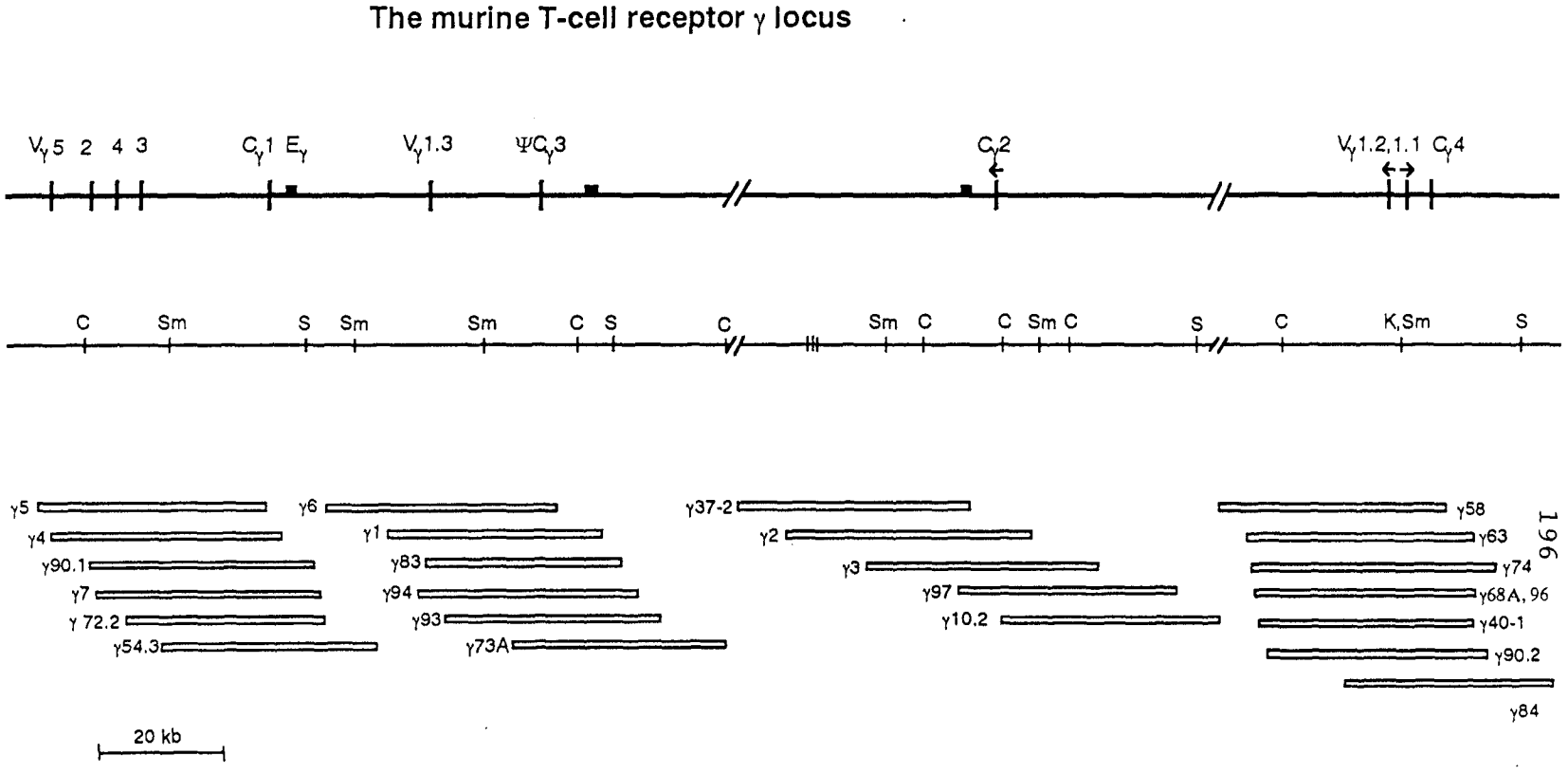


Figure 5. Preliminary map of the mouse TCR γ locus. Symbols as described in figures 1 and 2.

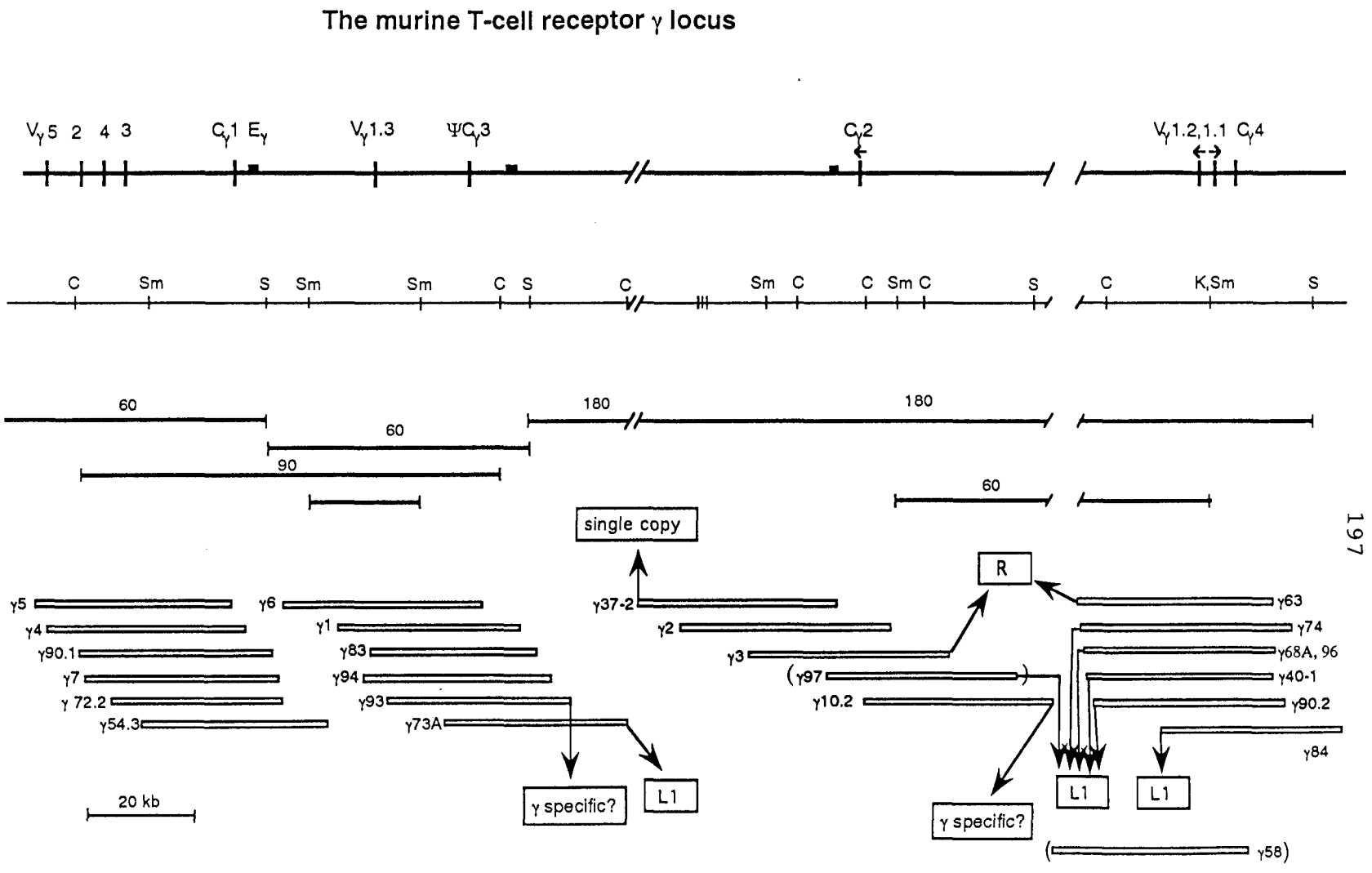


Figure 6. Preliminary map of the mouse TCR γ locus. See text for explanation. Symbols as described in figures 1 and 2.

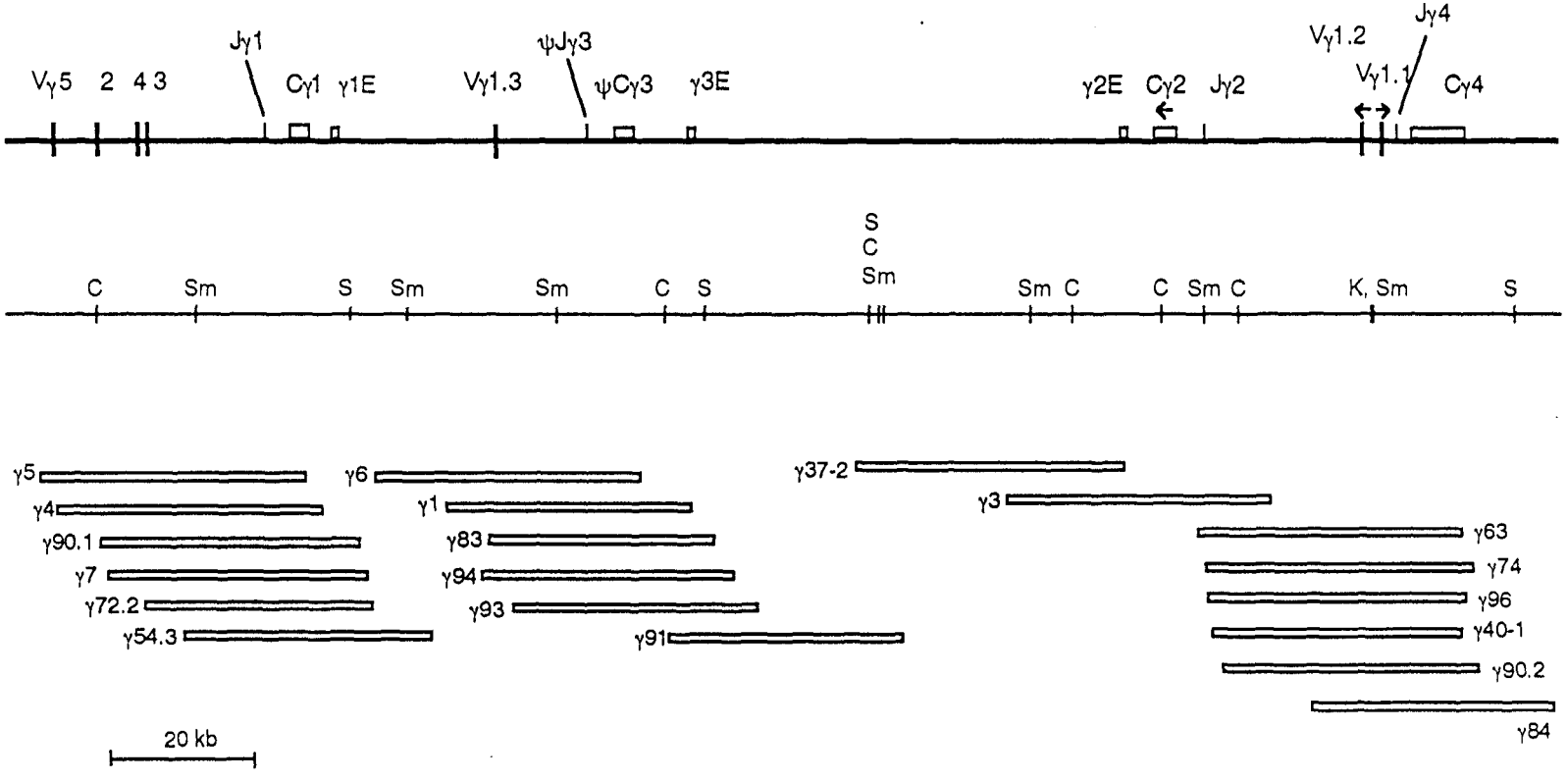


Figure 7. Final map of the TCR γ locus.