STRUCTURE AND FUNCTION OF THE MURINE T-CELL RECEPTOR GENES AND THE MURINE CLASS I GENES OF THE MAJOR HISTOCOMPATIBILITY COMPLEX

Thesis by

Astar Winoto

In partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1986

(submitted February 28, 1986)

To my lovely wife, Lulu and my sweet baby, Elena.

Acknowledgements

I'd like to thank my adviser, Leroy Hood, whose support and enthusiasm have provided me with a nice and warm environment for my graduate research. I'd also like to thank the following people in the laboratory, who have provided me with a stimulating environment, intellectually as well as socially, for the past four and a half years: Sandy Chang, Michael Steinmetz, Rick Barth, Mitch Kronenberg, Jerry Siu, Henry Sun and Shelley Mjolsness.

I am grateful to Bertha Jones, Connie Katz, Susan Mangrum and Cathy Elkins who have helped me tremendously in the preparation of this thesis and other manuscripts, Bertha Jones, Ria DeBruyn, Jessie Walker for running the laboratory smoothly, McCallum Foundation and the California Foundation for Biomedical Research for part of the financial support and last but not least, my parents, who make all of this possible.

Abstract

This thesis can be divided into three parts: Chapter Two and Chapter Three contain study on the class I genes of the major histocompatibility complex (MHC), Chapter Four and Chapter Five contain study on the T-cell receptor genes, and the appendices deal with my initial attempt to clone the cDNA encoding the T-cell receptor α chain and various other research projects that I have been involved in to some extent (mouse MHC class II genes, PDGF genes and rat class I and class II genes).

The first part of my thesis describes the study of the organization of the genes encoding the mouse class I MHC molecule. 54 cosmid clones containing 36 class I genes were isolated and, by restriction enzyme mapping, the 54 clones could be divided into 13 clusters. Using low-copy probes isolated from each cosmid cluster and the restriction enzyme site polymorphism of those probes, I was able to map each of the class I gene clusters into the precise location of the mouse MHC. Surprisingly, most of the class I genes map into the Tla region, only five class I genes (three cosmid clusters) map into the classical H-2 region. The functions of these class I genes in the Tla region are still largely unknown.

The remainder of my thesis contains the study on the T-cell receptor genes. In an effort to isolate the cDNA clone encoding the T-cell receptor α chain, I have isolated 64 T-cell specific cDNA clones, using a T-cell minus B-cell subtractive cDNA probe. The T-cell receptor α and β chain cDNA clones were among these 64 clones. Using the T-cell receptor α chain cDNA as a probe, I subsequently isolated clones encoding a germline variable(V) gene segment and cosmid clones spanning 120 kb of DNA encoding the joining(J) and constant(C) gene segments of the T-cell receptor α chain. Analysis of these clones, including sequencing of one germline V_α and six germline J_α gene segments, showed that the DNA recognition sequence for the α chain DNA rearrangment is similar to that of the β chain

counterpart. In contrast to the general J gene segment organization in the β chain, γ chain and the immunoglobulin gene families, I showed that the 18 J_α gene segments I analyzed were spread over 60 kb of DNA and lay as far as 63 kb 5' to the C_α gene.

In a step to dissect the structure function relationship of the T-cell receptor molecules, I have cloned and determined the nucleotide sequences of seven functional α chains and six β chains of the T-cell receptor genes from nine T-helper hybridomas specific for the C-terminal peptide of pigeon cytochrome c and the E class II molecule. Northern blot analyses using the isolated V_α and V_β gene segments were performed on the RNAs isolated from a total of 15 T-helper hybridomas specific for the C-terminal peptide of cytochrome c. A single V_α gene segment is predominantly used in these 15 T-helper hybridomas, whereas at least five different V_β gene segments are utilized. I conclude that the V_α gene segment is important for the cytochrome c response and might provide most of the contact residues with the C-terminal region of cytochrome c. I also found that the junctional sequences of the β chain may alter the antigen fine specificity of the T-cell clones. Finally, somatic hypermutation does not appear to play a crucial role in generating diversity for the T-cell receptor α or β chains.

TABLE OF CONTENTS

Chapter One

INTRODUCTION

A. MAJOR HISTOCOMPATIBILITY COMPLEX

1. Backgound

The major histocompatibility complex was first discovered as the locus encoding the antigen responsible for the tissue and tumor rejection in a transplantation experiment. As time goes on, it becomes obvious that the locus contains more than one gene; in fact, many genes involved in the immune response map to this locus. The major histocompatibility complex of the mouse (MHC) has been mapped to the chromosome 17. Using inbred strains of mice as well as congenic mice, which are genetically identical mice except for the genes of the MHC, immunogeneticists were able to identify two distinct complexes of MHC. The H-2 complex in the mouse encompassed 0.55 centimorgans and can be divided into four regions: the K region, the I region, the S region and the D region. The Tla complex encompassed around 1.5 centimorgans and can be divided into two regions: the Qa region and the Tla region (for review, see Klein, 1975).

The MHC loci encode three classes of gene families, denoted class I, class II and class III (Klein, 1975; Snell et al., 1976. Nathenson et al., 1981; Ploegh et al., 1981). The K and D regions contain the genes encoding the transplantation antigens K, D and L. They are class I molecules found on the cell surface of most nucleated cells and are highly polymorphic (Klein, 1979). More than 100 alleles have been defined at the K locus and similar number is also found for the D locus (Klein and Figueroa, 1981). The MHC alleles expressed in a given mouse strain comprise the haplotype of that strain. K^d, for example, denotes the K antigen or gene of the $H\text{-}2^d$ haplotye as seen in the BALB/c mouse. These molecules serve as restricting elements that permit cytotoxic T cells to recognize viral or tumor antigens on the surface of the infected or transformed cells. A phenomenon termed MHC restriction (Zinkernagel and Doherty, 1974). The I region can be subdivided further into five subregions by serological means: the I-A, I-B, I-J, I-E

and I-C subregions (Murphy, 1980). The I-A and I-E subregions encode the class II molecules which are expressed mainly on the lymphocytes and serve as restricting elements for the helper T cells. The I-J subregion was originally thought to encode the gene for a suppressor factor found in certain suppressor T cells. Subsequent genetic and molecular studies indicate that the I-J, as well as I-B and I-C subregions do not exist (Steinmetz et al., 1982b). The S region contains class III genes encoding certain components involved in the complement system (C2, Bf, Slp, C4) and a steroid biosynthesis enzyme, 21-hydrolase (Chaplin et al., 1983; Amor et al., 1985; White et al., 1984). The Tla and Qa regions comprised of the distal part of the MHC contain the Qa-2,3, Tla and Qa-1 genes, the proteins of which bear high level of homology to the K, D and L antigens and hence also called class I molecules. They are far less polymorphic than the K, D and L molecules and the functions of most of these molecules are still largely unknown (Flaherty, 1980).

The class I molecules are integral membrane glycoproteins with a molecular weight of 45 kilodaltons and for K, D, and L molecules, they are non-covalently associated with a 12 kilodalton polypeptide, β_2-microglobulin, to form the complete transplantation antigens. Protein sequence data suggest that the class I molecules are composed of three external domains, each about 90 amino acids in length, a transmembrane region of about 40 residues, and a cytoplasmic region of about 30 residues (Lopez de Castro et al., 1979; Coligan et al., 1981). Protein sequence comparison also indicates that the first two external domains are relatively variable and the third domain is highly conserved. It was suggested that the $\alpha 3$ domain is conserved because it interacts with the β_2-microglobulin. X-ray crystallography study of a human transplantation antigen shows that the $\alpha 1$ and $\alpha 2$ domains jointly form a structural domain and the $\alpha 3$ region indeed forms a domain associated with the β_2-microglobulin (Bjorkman et al., 1985).

Several groups have isolated and sequenced cDNA clones encoding the class I molecules (Ploegh et al., 1980; Bregegere et al., 1981; Kvist et al., 1981; Sood et al., 1981; Steinmetz et al., 1981a). Using the cDNA as a probe, it was shown that there are many class I genes. Sequencing of two of these genes (27.1 and L^dgenes) indicates that the genes contain eight exons that correlate precisely with the structural domains of class I antigens. Five discrete exons encode each for the leader domain, α1,α2, α3 domains, and the transmembrane domain; three exons encode the short cytoplasmic domain (Steinmetz et al., 1981b; Moore et al., 1982). A human class I gene sequence also shows the same exon intron structure as the mouse class I gene (Malissen et al., 1982).

2. Genetic organization of class I genes

In an effort to elucidate the genetic organization of the class I multigene family, I have helped in isolating 54 cosmid clones containing 36 class I genes from a Balb/c liver cosmid library. These clones could be divided into 13 clusters based on restriction enzyme maps (Chapter Two). To locate precisely the genetic map of those clusters, I have used low-copy probes generating from each cosmid cluster and using restriction enzyme polymorphism, I was able to map each of the cosmid cluster into the precise location of the mouse MHC. Surprisingly, most of the class I genes map into the Tla complex; only five class I genes (three clusters) map into the classical H-2 complex. Cluster six maps into either D or Qa regions, and it was shown later by molecular chromosomal walking work to map into the D region (Stephan et al., 1986). Hence, my work and others (Steinmetz et al., 1982a; Winoto et al., 1983; Fisher et al., 1985; Stephan et al., 1986) have established the following genetic organization of mouse class I genes in Balc/c mouse: two class I genes map in the K region, five class I genes map in the D region, eight class I genes map in the Qa 2 region, and 18 class I genes map in the Tla region.

B. T-CELL RECEPTOR

1. Background

The vertebrate immune system employs two kinds of antigen-specific receptors, immunoglobulin in B cells and T-cell receptor in T cells. The immunoglobulin molecules have been studied extensively; it is a heterodimer molecule composed of heavy and light chains. Each chain can be divided into a variable region and a constant region. Genomic DNA analysis indicates that the variable region is encoded by two or three gene segments: a variable (V) gene segment, a joining (J) gene segment, and for the heavy chain, a diversity (D) gene segment as well. The gene segments rearrange and join together during the lymphocyte differentiation to form a complete rearranged V gene. The RNA spicing will then remove the intron between J and C genes to form a mature transcript (for review, see Tonegawa, 1983).

The T-cell receptor molecules have been elusive for many years (for review see Kronenberg et al., 1983). The first breakthrough came with the generation of monoclonal antibodies that bound specifically to only one of a panel of T-cell clones (clonal specific antibody) by several laboratories (Allison et al., 1982; Haskins et al., 1983; Lancki et al., 1983; McIntyre and Allison, 1983; Meuer et al., 1983; Samelson et al., 1983). Using these monoclonal antibodies, it was shown that the antibodies affect antigen-specific activation; hence, the antibodies did bind to the antigen receptor. Immunoprecipitation studies using the monoclonals followed by two-dimensional gel analyses showed that the T-cell antigen receptors are composed of two subunits, α and β chains, each with molecular weight of 40-50 kilodaltons. Both chains appear to contain a variable region and a constant region (McIntyre and Allison, 1983; Kappler et al., 1983; Meuer et al., 1984).

The second breakthrough came when two groups independently isolated the cDNA clone encoding the T-cell receptor β chain by a differential and subtractive

screening approach (Hedrick et al., 1984; Yanagi et al., 1984). Mark Davis' group assumed several things in the identification of the T-cell receptor β chain cDNA: 1) the T-cell receptor molecule will be specific for the T cell only; 2) the T-cell receptor gene will rearrange during the T-cell differentiation analogous to the immunoglobulin gene; and 3) the message encoding the T-cell receptor will be quite abundant. The cDNA clone isolated has the above properties and by comparison with the protein sequence data generated from other laboratories, it was identified as the β chain of the T-cell receptor.

Progress in characterization of the T-cell receptor molecules have proceeded extremely rapidly since then. Another T cell-specific rearranging gene was isolated (γ chain gene, Saito et al., 1984a). Genomic analysis indicated that the variable regions of both β and γ genes are composed of several gene segments: V, D and J gene segments for the T-cell receptor β chain gene, and V and J gene segments for the γ chain gene. The α chain cDNA was not cloned until later (see below). Sequencing of the V and J gene segments of β and γ genes also revealed that the signals for DNA rearrangment in T-cell receptor are similar to that of immunoglobulin recognition sequences. The recognition signals consist of the conserved heptamer sequence CACAGTG and the A/T-rich nonamer sequence separated by either 11 bp (one-turn DNA helix) or 23 bp (two-turn DNA helix). The rearrangment can presumably occur only between a gene segment with a one-turn DNA recognition sequence and a gene segment with a two-turn DNA recognition sequence. The V gene segments of β and γ chain genes contain a two-turn recognition sequence, the J gene segments contain a one-turn recognition sequence, and the D gene segment in β chain gene has a one-turn recognition sequence on the 5' side and a two-turn recognition sequence on the 3' side of the coding region. Hence, in β chain gene, the V gene segment can join directly to the J gene segment (for review, see Kronenberg et al., 1986).

The genomic organization of the mouse β chain and γ chain have been characterized extensively (Chien et al., 1984a; Gascoigne et al., 1984; Kavaler et al., 1984; Malissen et al., 1984; Siu et al., 1984; Hayday et al., 1985). The gene encoding the T-cell receptor β chain is composed of 20-30 V gene segments (Barth et al., 1985; Behlke et al., 1985), two D gene segments (Kavaler et al., 1984; Siu et al., 1984), and six functional J gene segments each in front of two C genes (Chien et al., 1984a; Gascoigne et al., 1984; Malissen et al., 1984). The gene encoding the γ chain is composed of yet unkown but limited number of V gene segments, and one J gene segment each in front of four C genes.

The T-cell receptor genes share several strategies for diversification as the immunoglobulin genes: 1) junctional variation—the flexibility of the sites at which the gene segments join together, deleting different nucleotides from the V, D or J gene segments leading to codon changes at the junctions of these rearranged gene segments; 2) N-region diversity—the random addition of nucleotides at both sides of the rearranging gene segments; 3) multiplicities of germline gene segments—there are many V gene segments in T-cell receptor as well as immunoglobulin genes, four J gene segments each in the heavy chain, κ chain, λ chain of the immunoglobulin, 12 functional J gene segments in the β chain gene and four J gene segments in the γ chain gene; and 4) combinatorial association—any V gene segments can join to any downstream D or J gene segments. Somatic hypermutation, which can alter up to 3% of a given V gene sequences, has not been seen in the β chain or γ chain genes.

T-cell recognition differs from the B-cell recognition in that the T-cell receptor has to recognize antigen in conjunction with the MHC gene products (MHC restriction) (Kindred and Shreffler, 1972; Katz et al., 1973; Rosenthal and Shevach, 1973; Zinkernagel and Doherty, 1974). During ontogeny, the T cells were "educated" in the thymus such that they will recognize only antigens in the

context of self MHC products. A phenomenon termed thymus education (Weissman, 1967; Cantor and Weissman, 1976). These phenomena (MHC restriction and thymus education) raise questions on how the T-cell receptor recognizes the antigen/MHC and how thymus education works. Several models have been proposed as to how the T-cell receptor recognizes the antigen/MHC together. The single receptor hypothesis proposes that both antigen and MHC molecules are bound by a single receptor (Zinkernagel and Doherty, 1974; Cohen and Eisen, 1977; Matzinger, 1981; Schrader, 1982). The dual receptor hypothesis proposes that two distinct recognition elements bind, respectively, antigen and MHC molecule (Blanden and Ada, 1978; Cohn and Epstein, 1978; von Boehmer et al., 1978; Parham, 1984). A variety of experiments and analyses of the secondary structure of the T-cell receptor V regions suggest that the T-cell receptor is just a single molecular complex (Kappler et al., 1981; see Kronenberg et al., 1986 for review). A new version of the dual receptor model suggests that distinct portion of the T-cell receptor may bind antigen and MHC molecule, hence, the α chain of the T-cell receptor may bind the antigen and the β chain may bind the MHC molecule or vice versa (Patten et al., 1984; Pernis and Axel, 1985). Thymus education, on the other hand, has been proposed to involve the selective destruction of the T cells that bind to the self MHC molecule with high affinity and the selective proliferation of the T cells that bind to the self MHC molecule loosely (Marrack and Kappler, 1986). Another model proposes the γ chain molecule in association with the β chain molecule to recognize the self MHC molecule during development; when the thymus matures, the γ molecule is slowly replaced by the α chain molecule that, in association with the β chain molecule, now recognize the antigen/MHC molecule (Raulet et al., 1985; Pernis and Axel, 1985).

2. The gene encoding the T-cell receptor α chain

When I started the project on T-cell receptor in March 1984, my first

goal was to isolate the cDNA clone encoding the T-cell receptor α chain. I made the same assumptions as Mark Davis had done before in isolating the T-cell receptor β chain: that the clone will be T cell-specific and it will show rearrangment in a genomic blot analyses. Using two subtractive probes of T cell cDNA minus B cell RNA and B cell cDNA minus macrophage RNA, I was able to isolate 64 T cell-specific clones from a T-cell hybridoma V1.9.2 cDNA library. The clones were characterized for the ability to rearrange during the T-cell differentiation. Southern genomic blot analyses were performed using the inserts isolated from the 64 λ clones. The germline mouse DNA, BW5147 fusion partner DNA, and the T-cell hybridoma V1.9.2 DNA digested by four different enzymes were included in the Southern blot analyses. Only one clone showed DNA rearrangment; this clone, however, was a T-cell receptor β chain cDNA. When Mark Davis' group and Tonegawa's group identified the T-cell receptor α chain (Chien et al., 1984b; Saito et al., 1984b), I used oligonucleotide specific for the α chain in a hybridization to those 64 T cell-specific clones. One of them did contain the constant region of the T-cell receptor α chain. The reason that I did not detect the DNA rearrangment using a constant region probe became apparent when I anayzed the genomic organization of the α chain (see below). The J gene segment that the T-cell hybridoma V1.9.2 uses is located 21 kb 5' to the C_α gene. Hence, a constant region probe alone cannot detect the rearrangment of the V gene segment into the J_α28 used by the hybridoma V1.9.2.

Using the T-cell receptor α chain cDNA as a probe, I subsequently isolated clones encoding a germline variable (V) gene segment and cosmid clones spanning 120 kb of DNA encoding the joining (J) and constant (C) gene segments of the T-cell receptor α chain. Analysis of these clones including sequencing of one germline V_α and six germline J_α gene segments showed that the DNA recognition sequence for the α chain DNA rearrangment is similar to that of the β chain

counterpart. The gene encoding the α-chain also employs the diversification strategies similar to that of the β and γ chain genes. In contrast to the general J gene segment organization in the β chain, γ chain and the immunoglobulin gene families, I showed that the 18 J_α gene segments I analyzed were spread over 60 kb of DNA and lie as far as 63 kb 5' to the C_α gene (Winoto et al., 1985).

3. Pigeon cytochrome c/I-E specific T-cell receptors

In a step to dissect the structure function relationship of the T-cell receptor molecules, I have cloned and determined the nucleotide sequences of seven functional α chains and six β chains of the T-cell receptor genes from nine T-helper hybridomas specific for the C-terminal peptide of pigeon cytochrome c and the I-E class II molecule. Northern blot analyses using the isolated V_α and V_β gene segments were performed on the RNAs isolated from a total of 15 T-helper hybridomas specific for the C-terminal peptide of cytochrome c. A single V_α gene segment is predominantly used in these 15 T-helper hybridomas, whereas at least five different V_β gene segments are utilized. I conclude that the V_α gene segment is important for the cytochrome c response and might provide most of the contact residues with the C-terminal region of cytochrome c. I also found that the junctional sequences of the β chain may alter the antigen fine specificity of the T-cell clones. The patterns of T-cell gene segment usage and alterations in antigen specificity associated with changes in the third hypervariable region are similar to those seen in B-cell immunoglobulin responses to antigen. These exclude a simple dual site-single receptor model of T-cell recognition as mentioned above and are consistent with the idea that the T-cell receptor and immunoglobulin molecules are similar in structure to one another. Finally, somatic hypermutation does not appear to play a crucial role in generating diversity for the T-cell receptor α or β chains.

REFERENCES

Allison, J. P., McIntyre, B. W., and Bloch, D. (1982). Tumor-specific antigen of murine T lymphoma defined with monoclonal antibody. *J. Immunol.* **129,** 2293-2300.

Amor, M., Tosi, M., Duponchel, C., Steinmetz, M., and Meo, T. (1985). Liver mRNA probes disclose two cytochrome P-450 genes duplicated in tandem with the complement C4 loci of the mouse H-2 S region. *Proc. Natl. Acad. Sci. USA* **82,** 4453-4457.

Barth, R., Kim, B., Lan, N., Hunkapiller, T., Sobieck, N., Winoto, A., Gershenfeld, H., Okada, C., Hansburg, D., Weissman, I., and Hood, L. (1985). The murine T-cell receptor employs a limited repertoire of expressed V_β gene segments. *Nature* **316,** 517-523.

Behlke, M. A., Spinella, D. G., Chou, H., Sha, W., Hartl, D. L., and Loh, D. Y. (1985). T-cell receptor β chain expression dependence on relative few variable genes. *Science* **229,** 566-570.

Bjorkman, P. J., Strominger, J. L., and Wiley, D. C. (1985). Crystallization and X-ray diffraction studies on the histocompatibility antigens HLA-A2 and HLA-A28 from human cell membranes. *J. Mol. Biol.* **186,** 205-210.

Blanden, R. V., and Ada., G. L. (1978). A dual recognition model for cytotoxic T cells based on thymic selection of precursor with low affinity for self H-2 antigens. *Scand. J. Immunol.* **7,** 181-190.

Bregegere, F., Abastado, J. P., Kvist, S., Rask, L., Lalanne, J. L., Garoff, H., Cami, B., Wiman, K., Larhammar, D., Peterson, P. A., Gachelin, G., Kourilsky, P., and Dobberstein, B. (1981). Structure of C-terminal half of two H-2 antigens from cloned mRNA. *Nature* **292,** 78-81.

Cantor, H., and Weissman, I. (1976). Development and function of subpopulations of thymocytes and T lymphocytes. In *Progress in Allergy*, Vol. 20, pp. 1-64.

Chaplin, D. D., Woods, D. E. Whitehead, A. S., Goldberger, G., Colten, H. R., and Seidman, J. G. (1983). Molecular map of the murine S region. *Proc. Natl. Acad. Sci. USA* **80**, 6947-6951.

Chien, Y.-H., Gascoigne, N. R. J., Kavaler, J., Lee, N. E., and Davis, M. M. (1984a). Somatic recombination in a murine T-cell receptor gene. *Nature* **309**, 322-326.

Chien, Y., Becker, D., Lindsten, R., Okamura, M., Cohen, D., and Davis, M. (1984b). A third type of murine T-cell receptor gene. *Nature* **312**, 31-35.

Cohen, R. J., and Eisen, H. (1977). Interactions of macromolecules on cell membranes and restriction of T-cell specificity by products of the major histocompatibility complex. *Cell. Immunol.* **32**, 1-9.

Cohn, M., and Epstein, R. (1978). T-cell inhibition of humoral responsiveness, II. Theory on the role of restrictive recognition in immune regulation. *Cell Immunol.* **39**, 125-153.

Coligan, J. E., Kindt, T. J., Uehara, H., Martinko, J., and Nathenson, S. G. (1981). Primary structure of a murine transplantation antigen. *Nature* **291**, 35-39.

Fisher, D. A., Hunt, S. W. III, and Hood, L. (1985). Structure of a gene encoding a murine thymus leukemia antigen, and organization of Tla genes in the Balb/c mouse. *J. Exp. Med.* **162**, 528-545.

Flaherty, L. (1980). Tla region antigens. In *The Role of the Major Histocompatibilty Complex in Immunobiology*, ed. M. Dorf, (Garland, New York), pp. 33-57.

Gascoigne, N. R. J., Chien, Y.-H., Becker, D. M., Kavaler, J., and Davis, M. M. (1984). Genomic organization and sequence of T-cell receptor β-chain constant-and joining-region genes. *Nature* **310**, 387-391.

Haskins, K., Kubo, R., White, J., Pigeon, M., Kappler, J., and Marrack, P. (1983). The major histocompatibility complex-restricted antigen receptor on T cells. I. Isolation with a monoclonal antibody. *J. Exp. Med.* **157**, 1149-1169.

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

Hedrick, S. M., Cohen, D. I., Nielsen, E. A., and Davis, M. M. (1984). Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. *Nature* **308**, 149-153.

Kappler, J. W., Skidmore, B., White, J., and Marrack, P. (1981). Antigen-inducible, H-2 restricted interleukin-2-producing T-cell hybridomas. Lack of independent antigen and H-2 recognition. *J. Exp. Med.* **153**, 1198-1214.

Kappler, J., Kubo, R., Haskins, K., Hannum, C., Marrack, P., Pigeon, M., McIntyre, B., Allison, J., and Towbridge, I. (1983). The major histocompatibility complex-restricted antigen receptor on T cells in mouse and man: identification of constant and variable peptides. *Cell* **35**, 295-302.

Katz, D. H., Hamaoka, T., and Benacerraf, B. (1973). Cell interactions between histoincompatible T and B lymphocytes. II. Failure of physiologic cooperative interactions between T and B lymphocytes from allogeneic donor strains in humoral response to hapten-protein conjugates. *J. Exp. Med.* **137**, 1405-1418.

Kavaler, J., Davis, M. M., and Chien, Y.-H. (1984). Localization of a T-cell receptor diversity region element. *Nature* **310**, 421-423.

Kindred, B., and Shreffler, D. C. (1972). H-2 dependence of cooperation between T and B cells in vivo. *J. Immunol.* **109**, 940-943.

Klein, J. (1975). Biology of the Mouse Histocompatibility-2 Complex. Springer-Verlag, New York.

Klein, J. (1979). The major histocompatibility complex of the mouse. *Science* **203**, 516-521.

Klein, J., and Figueroa, F. (1981). Polymorphism of the mouse H-2 loci. *Immunol. Rev.* **60**, 23-57.

Kronenberg, M., Kraig, E., and Hood, L. (1983). Finding the T-cell antigen receptor: Past attempts and future promise. *Cell* (minireview) **34**, 327-329.

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.

Kvist, S., Bregegere, F., Rask, L., Cami, B., Garoff, H., Daniel, F., Wiman, K., Larhammar, D., Abastado, J. P., Gochelin, G., Peterson, P. A., Dobberstein, B., and Kourilsky, P. (1981). cDNA clone coding for part of a mouse H-2^d major histocompatiblity antigen. *Proc. Natl. Acad. Sci. USA* **78**, 2772-2776.

Lancki, D. W., Lorber, M. I., Loken, M. R., and Fitch, F. W. (1983). A clone-specific monoclonal antibody that inhibits cytolysis of a cytolytic T-cell clone. *J. Exp. Med.* **157**, 921-935.

Lopez de Castro, J. A., Orr, H. T., Robb, R. J., Kostyk, T. G., Mann, D. L., and Strominger, J. L. (1979). Complete amino acid sequence of a papain-solubilized human histocompatibility antigen, HLA-B7. I. Isolation and amino acid composition of fragments and of tryptic and chymotryptic peptides. *Biochemistry* **18**, 5704-5711.

Malissen, M., Malissen, B., and Jordan, B. R. (1982). Exon-intron organization and complete nucleotide sequence of an HLA gene. *Proc. Natl. Acad. Sci. USA* **79**, 520-524.

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

Marrack, P., and Kappler, J. (1986). The T cell and its receptor. *Sci. Am.* **254**, 36-45.

Matzinger, P. (1981). A one-receptor view of T-cell behavior. *Nature* **292**, 497-501.

McIntyre, B. W., and Allison, J. P. (1983). The mouse T-cell receptor: structural heterogeneity of molecules of normal T cells defined by xenoantiserum. *Cell* **34**, 739-746.

Meuer, S. C., Fitzgerald, K. A., Hussey, R. E., Hodgdon, J. C., Schlossman, S. F., and Reinherz, E. L. (1983). Clonotypic structures involved in antigen-specific human T-cell function. Relationship to the T3 molecular complex. *J. Exp. Med.* **157**, 705-719.

Meuer, S. C., Acuto, O., Hercend, T., Schlossman, S. F., and Reinherz, E. L. (1984). The human T-cell receptor. *Ann. Rev. Immunol.* **2**, 23-50.

Moore, K. W., Sher, B. T., Sun, Y. H. Eakle, K. A., and Hood, L. (1982). DNA sequence of a gene encoding a Balb/c mouse L^d transplantation antigen. *Science* **215**, 679-682.

Murphy, D. B. (1980). Genetic fine structure of the H-2 gene complex. In *The Role of the Major Histocompatibility Complex in Immunology*, ed. M. Dorf, (Garland, New York), pp. 1-32.

Nathenson, S. G., Uehara, H., Ewenstein, B. M., Kindt, T. J., and Coligan, J. E. (1981). Primary structural analysis of transplantation antigens of the murine H-2 major histocompatibility complex. *Ann. Rev. Biochem.* **50**, 1025-1052.

Parham, P. (1984). A repulsive view of MHC restriction. *Immunol. Today* **5**, 89-92.

Patten, P., Yokota, T., Rothbard, J., Chien, Y.-H., Arai, K.-I., and Davis, M. M. (1984). Structure, expression and divergence of T-cell receptor β chain variable regions. *Nature* **312**, 40-46.

Pernis, B., and Axel, R. (1985). A one and a half receptor model for MHC restriction antigen recognition by T lymphocytes. *Cell* **41**, 13-16.

Ploegh, H. L., Orr, H. T., and Strominger, J. L. (1980). Molecular cloning of a human histocompatibility antigen cDNA fragment. *Proc. Natl. Acad. Sci. USA* **77**, 6081-6085.

Ploegh, H. L., Orr, H. T., and Strominger, J. L. (1981). Major histocompatibility antigens: the human (HLA-A, -B, -C) and murine (H-2K, H-2D) class I molecules. *Cell* **24**, 287-299.

Raulet, D. H., Garman, R. D., Saito, H., and Tonegawa, S. (1985). Segments of the murine T-cell antigen receptor. *Nature* **311**, 344-356.

Rosenthal, A. S., and Shevach, E. M. (1973). Function of macrophages in antigen recognition by guinea pig T lymphocytes. *J. Exp. Med.* **138**, 1194-1212.

Saito, H., Kranz, D. M., Takagaki, Y., Hayday, A. C., Eisen, H. N., and Tonegawa, S. (1984a). Complete primary structure of a heterodimeric T-cell receptor deduced from cDNA sequences. *Nature* **309**, 757-762.

Saito, H., Kranz, D., Takagaki, Y., Hayday, A., Eisen, H., and Tonegawa, S. (1984b). A third rearranged and expressed gene in a clone of cytotoxic T lymphocytes. *Nature* **312**, 36-40.

Samelson, L. E., Germain, R. N., and Schwartz, R. H. (1983). Monoclonal antibodies against the antigen receptor on a cloned T-cell hybrid. *Proc. Natl. Acad. Sci. USA* **80**, 6972-6976.

Schrader, J. (1982). A single T-cell receptor: A speculative review of the intrathymic generation and modulation of the repertoire. *Thymus* **4**, 181-207.

Siu, G., Kronenberg, M., Strauss, E., Haars, R., Mak, T., and Hood, L. (1984). The structure, rearrangement and expression of D_β gene segments of the murine T-cell antigen receptor. *Nature* **311**, 344-356.

Snell, G. D., Dausset, J., and Nathenson, S. (1976). *Histocompatibility*. New York: Academic Press.

Sood, A. K., Pereira, D., and Weissman, S. M. (1981). Isolation and partial nucleotide sequence of a cDNA clone for human histocompatibility antigen HLA-B by use of an oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* **78**, 616-620.

Steinmetz, M., Frelinger, J. G., Fisher, D., Hunkapiller, T., Pereira, D., Weissman, S. M., Uehara, H., Nathenson, S., and Hood, L. (1981a). Three cDNA clones encoding mouse transplantation antigens: homology to immunoglobulin genes. *Cell* **24**, 125-134.

Steinmetz, M., Moore, K. W., Frelinger, J. G., Sher, B. T., Shen, F.-W., Boyse, E. A., and Hood, L. (1981b). A pseudogene homologous to mouse transplantation antigens: transplantation antigens are encoded by eight exons that correlate with protein domains. *Cell* **25**, 683-692.

Steinmetz, M., Winoto, A., Minard, K., and Hood, L. (1982a). Clusters of genes encoding mouse transplantation antigens. *Cell* **28**, 489-498.

Steinmetz, M., Minard, K., Horvath, S., McNicholas, J., Frelinger, J., Wake, C., Long, E., Mach, B., and Hood, L. (1982b). A molecular map of the immune response region from the major histocompatiblity complex of the mouse. *Nature* **300**, 35-42.

Stephan, D., Sun, H., Lindahl, K. F., Meyer, E., Hammerling, G., Hood, L., and Steinmetz, M. (1986). Organization and evolution of D region class I genes in the mouse MHC. *J. Exp. Med.*, in press.

Tonegawa, S. (1983). Somatic generation of antibody diversity. Nature **302**, 575-581.

von Boehmer, H., Haas, W., and Jerne, N. K. (1978). Major histocompatibility complex-linked immune-responsiveness is acquired by lymphocytes of low responder mice differentiating in thymus of high-responder mice. *Proc. Natl. Acad. Sci. USA* **75**, 2439-2442.

Weissman, I. L. (1967). Thymus cell migration. *J. Exp. Med.* **126**, 291-304.

White, P. C. Chaplin, D. D., Weis, J. H., Dupont, B., New, M. I., and Seidman, J. G. (1984). Two steroid 21-hydroxylase genes are located in the murine S region. *Nature* **312**, 465-467.

Winoto, A., Steinmetz, M., and Hood, L. (1983). Genetic mapping in the major histocompatibility complex by restriction enzyme polymorphisms: Most mouse class I genes map to the Tla complex. *Proc. Natl. Acad. Sci. USA* **80**, 3425-3429.

Winoto, A., Mjolsness, S., and Hood, L. (1985). Genomic organization of the genes encoding mouse T-cell receptor α chain. *Nature* **316**, 832-836.

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

Zinkernagel, R. M., and Doherty, P. C. (1974). Restriction of *in vitro* T-cell mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. *Nature* **248**, 701-702.

Chapter Two

CLUSTERS OF GENES ENCODING MOUSE TRANSPLANTATION ANTIGENS

Published in

Cell **28,** 489-498, 1982

Cell, Vol. 28, 489–498, March 1982, Copyright © 1982 by MIT

Clusters of Genes Encoding Mouse Transplantation Antigens

Michael Steinmetz, Astar Winoto, Karyl Minard and Leroy Hood
Division of Biology
California Institute of Technology
Pasadena, California 91125

Summary

We constructed a cosmid library from BALB/c mouse sperm DNA and isolated 64 cosmid clones with cDNA probes for transplantation antigens (class I molecules). Of these clones, 54 mapped into 13 gene clusters containing 36 distinct class I genes and encompassing 837 kilobases of DNA. One gene cluster mapped to the L region and a second cluster with seven genes to the Qa-2,3 region of the major histocompatibility complex. Restriction map and Southern blot analyses suggest that there are subgroups of class I genes. Using a 5′ flanking sequence of the L gene as a hybridization probe, we show the L gene to be present in mouse strains expressing this antigen but deleted or mutated in strains failing to express it. Our data suggest that gene duplication and deletion presumably by homologous but unequal crossing-over has altered the size and organization of the class I clusters in different mouse strains and probably is an important mechanism for generating polymorphism in these genes. Analysis of the 36 class I genes with cDNA probes specific for the 5′ and 3′ ends shows that the exon encoding the third external domain is far more conserved than those encoding the first and second external domains of the transplantation antigen. These differences in variability have interesting functional implications.

Introduction

The major histocompatibility complex (MHC), which encodes at least three classes of genes, plays a critical role in regulating the immune response (Klein, 1975; Zinkernagel and Doherty, 1980; Ploegh et al., 1981). The class I genes encode transplantation antigens that are required for the lysis of a virally infected or neoplastically transformed cell by cytotoxic T cells (T-cell immunosurveillance). The class II genes encode cell-surface molecules on bone-marrow-derived cells (B cells, T cells, macrophages) that are involved in lymphocyte communication and the induction of immune responses. The class III genes encode several complement components. We have studied the class I genes and antigens because they constitute an intriguing model system for cell–cell recognition.

Transplantation antigens are found on virtually all somatic cells of mammalian organisms. They consist of a 45,000 dalton integral membrane protein (class I molecule) noncovalently associated with the 12,000 dalton polypeptide, β2-microglobulin, not encoded by the MHC. Amino acid sequence data suggest that the class I molecule has three external domains, each about 90 residues in length, a transmembrane piece of about 40 residues and a cytoplasmic domain of approximately 30 residues (Lopez de Castro et al., 1979; Coligan et al., 1981; Nathenson et al., 1981).

In the BALB/c mouse, four serologically defined transplantation antigens, K, D, L and R (Hansen et al., 1981), encoded by genes that map at either end of the MHC, flank the class II and class III genes (Figure 1). The class I genes are extremely polymorphic, with more than 100 alleles having been defined at the K locus and 100 alleles at the D locus. The MHC or H-2 alleles expressed in a given mouse strain comprise the haplotype of that strain; for example, K^d denotes a K antigen or gene of the H-2^d haplotype as seen in the BALB/c mouse. Adjacent to the MHC on mouse chromosome 17 is the Tla complex, which encodes the Qa and TL antigens (Figure 1). It has been proposed that these are class I molecules based on their similarities to the classical transplantation antigens in their size, peptide map profiles and association with β2-microglobulin (Michaelson et al., 1977; Flaherty, 1980; Stanton and Hood, 1980; Soloski et al., 1981).

Several investigators have isolated and sequenced cDNA clones encoding class I molecules (Ploegh et al., 1980; Brégégère et al., 1981; Kvist et al., 1981; Sood et al., 1981; Steinmetz et al., 1981a). Using class I cDNA clones as probes on Southern blots of germline or liver DNA from several inbred strains of mice, we have demonstrated that 10 to 15 bands hybridize with the class I probes (Steinmetz et al., 1981a, 1981b, also see Cami et al., 1981). We have isolated 30 to 40 distinct clones from a BALB/c sperm DNA library constructed in phage λ using the cDNA probes (Steinmetz et al., 1981b). Two class I genes in these clones have been completely sequenced. The first, gene 27.1, is a pseudogene that maps to the Qa-2,3 region (Figure 1) (Steinmetz et al., 1981b). Transformation and sequencing studies demonstrated that the second, gene 27.5, encodes an L^d molecule (Goodenow et al., 1982; Moore et al., 1982). Both genes contain eight exons that correlate precisely with the structural domains of transplantation antigens—discrete exons are found for the signal peptide, for each of the three external domains and for the transmembrane domain, and three exons encode the short cytoplasmic domain. A human class I gene recently sequenced shows precisely the same exon–intron organization except that it contains only two cytoplasmic exons (Malissen et al., 1982).

We have recently constructed a cosmid genomic library containing large inserts (32 to 46 kilobases [kb]) of BALB/c sperm DNA. We use cosmid clones to demonstrate that there are at least 36 class I genes,

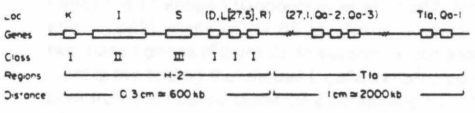


Figure 1. Map of the Major Histocompatibility or H-2 Complex on Chromosome 17 of the BALB/c Mouse

The genetic map includes the class I pseudogene 27.1 located in the Qa-2,3 region (Steinmetz et al., 1981b) and the L^d gene 27.5 (Goodenow et al., 1982; Moore et al., 1982). The order of loci included in brackets is not known. For the conversion of centiMorgans (cM) to kilobase pairs (kb) see text.

which fall into 13 discrete clusters encompassing 837 kb of DNA in the BALB/c mouse genome.

Results and Discussion

Cosmid Clones Contain at Least 36 Class I Genes Encompassing 837 Kilobases of DNA

A cosmid library was constructed from BALB/c sperm DNA by the techniques of Grosveld et al. (1981) and Ish-Horowicz and Burke (1981). cDNA probes specific for the 5′ and 3′ ends of class I genes (Figure 2 and Experimental Procedures) were used to screen an unamplified genomic library of about 250,000 cosmid clones (~3 genome equivalents). A total of 64 cosmid clones was selected with these probes.

The selected clones were mapped by digestion with ten different restriction enzymes (Sal I, Cla I, Sma I, Sac II, Xho I, Xma III, Nru I, Kpn I, Hpa I and Bam HI). The first seven of these restriction enzymes have one or two CpG dinucleotides in their recognition sequences. This dinucleotide is underrepresented in eucaryotic DNA, and accordingly, these enzymes generally cleave in only a few sites in the eucaryotic inserts of the cosmid DNA.

Of the 64 cosmid clones, 50 demonstrated partial restriction map overlaps, which allowed us to define 9 different clusters of class I genes containing from 2 to 16 overlapping cosmid clones. In addition, four clones did not overlap with one another or with the nine clusters of class I genes. The restriction maps for the inserts in these 54 cosmid clones are given in Figure 3 and are ordered into the 13 clusters, which range in size from 35 to 191 kb and contain between one and seven class I genes. In total, these 13 clusters encompass 837 kb of DNA with 36 class I genes (Figure 4).

The size of the inserts in the 54 cosmid clones ranged between 32 and 46 kb, with an average size of 38 kb. Two of the 54 clones (52.2 in cluster 6 and 36.2 in cluster 9) probably contain deletions that occurred after the construction of the cosmid library because their inserts were 31 and 28 kb, respectively—below the minimum length of 32 kb required for the successful packaging of the cosmid vector.

Of the ten remaining cosmid clones (64 minus 54), five have not been mapped because their restriction

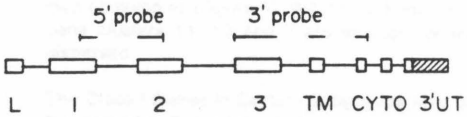


Figure 2. A Diagram of the 5′ and 3′ Class I cDNA Probes Used for Isolation and Analysis of the Cosmid Clones

The sequences contained in the 5′ and 3′ class I probes are depicted above the diagram for a class I gene with individual exons for the leader or signal peptide (L), the three external domains (1, 2, 3), the transmembrane domain (TM), and three exons for the cytoplasmic domain (CYTO) and the 3′ untranslated region (3′ UT, shown as hatched box). The 5′ cDNA probe extends from codon 63 to 160, and the 3′ cDNA probe extends from codon 183 to 329. These probes were derived from cDNA clones pH-2IIa and pH-2III (see Experimental Procedures).

map patterns were complicated, perhaps as a result of contaminating DNA fragments derived by deletions of the inserted DNAs. One clone contained an insert of only 18 kb and did not overlap with any other clone. The remaining four cosmid clones showed partial overlaps to certain clones in the clusters, but did not fit unambiguously into any of the clusters. Thus these clones may have been generated by ligation of smaller noncontiguous DNA fragments. Alternatively, they may represent additional duplicated and partially homologous class I genes.

In Table 1 we have correlated the 36 class I genes contained in our 13 clusters to their respective Bam HI fragments detected by the 3′ cDNA probe. This correlation will allow the easy identification of class I genes for which polymorphic restriction patterns are observed in Southern blot analyses of different mouse inbred strains.

Two Class I Clusters Map to the L and Qa Regions

The λ clones containing the two completely sequenced class I genes, 27.1 (Steinmetz et al., 1981b) and 27.5 (Moore et al., 1982), have been located in two distinct class I gene clusters by comparison of their restriction maps to those of the corresponding cosmid clones. The pseudogene 27.1 maps to the 191 kb cluster 1 containing seven distinct class I genes (Figure 3). Since gene 27.1 has been mapped to the Qa-2,3 region of the BALB/c H-2 complex (Steinmetz et al., 1981b), this gene cluster also must map into the Qa-2,3 region. We confirmed that cluster 1 maps to the Qa-2,3 region by Southern blot analyses of congenic mice strains, using a low-copy-number fragment isolated from the 3′ flanking region of gene 2 (see below). The 15 kb Kpn fragment from cluster I detected by this probe is present in B6-H-2^k and B6.K1 mice and is absent in B6.K2 and B6-Tla^a mice. These pairs of recombinant congenic mice differ only in their Qa-2,3 regions (see Steinmetz et al., 1981b). Thus these data allow us to map the class I cluster 1 to the Qa-2,3 region unequivocally.

From restriction analysis, gene 27.5, which en-

codes the L^d antigen (Goodenow et al., 1982; Moore et al., 1982), maps to the 68 kb cluster 2 containing two class I genes (Figure 3). In support of our assignment is the finding that mouse L cells transfected with DNA from the cosmid clone 59.2 containing a putative L^d gene react with two monoclonal antibodies to L^d antigens (Goodenow et al., 1982; R. S. Goodenow, unpublished results). We do not know whether the other 11 gene clusters map to the H-2 complex of mouse chromosome 17. Indeed, with the demonstration that members of the α-globin-gene family can map to different chromosomes (Leder et al., 1981), there is a distinct possibility that one or more of the gene clusters described above is not encoded on chromosome 17.

Cluster I Has Seven Class I Genes and Contains 191 Kilobases of DNA

Sixteen cosmid clones map into cluster 1, which encompasses 191 kb of DNA (Figure 3). The seven genes found in cluster 1 display extensive restriction site homology (Figure 5). The genes at the 3′ end of the cluster and their flanking sequences are almost identical by restriction mapping (also see Table 1). This homology suggests that these genes arose by relatively recent duplication, presumably through homologous but unequal crossing-over. The class I genes at the 5′ end of this cluster are homologous to one another, but distinctly less homologous to those genes at the 3′ end of the cluster. These data suggest that there have been multiple and recent gene duplication events in this gene cluster. A determination of the location of the 5′ and 3′ ends of these class I genes with the 5′ and 3′ cDNA probes demonstrates that all seven genes are transcribed from the same DNA strand (Figure 5).

Certain Gene Clusters May Contain Subgroups of Closely Related Class I Genes

The appearance of strong bands in Southern blot analyses corresponding to multiple class I genes is striking. For instance, the strong bands at 3.1 kb, 2.7 kb and 2.6 kb each correspond to at least three class I genes (Table 1). In several cases these related genes are linked in clusters (Table 1). This raises the possibility that subgroups of class I genes might be contained in closely linked gene clusters. Future experiments with subgroup-specific probes (see below) will allow a more detailed analysis of these subgroups of class I genes and might help to clarify the linkage relationship between members of different subgroups.

In certain cases, closely related class I genes may be dispersed throughout the H-2 complex. Striking homologies in the restriction map patterns are found among the class I genes in distint clusters. For example, gene 1 of the cosmid clone 17.1 (cluster 11), the gene found in the cosmid clone 18.1 (cluster 13) and the L gene (gene 2 in cluster 2) show restriction map homologies (Figure 3). We do not know whether gene clusters 11, 13 and 2 are adjacent or widely dispersed.

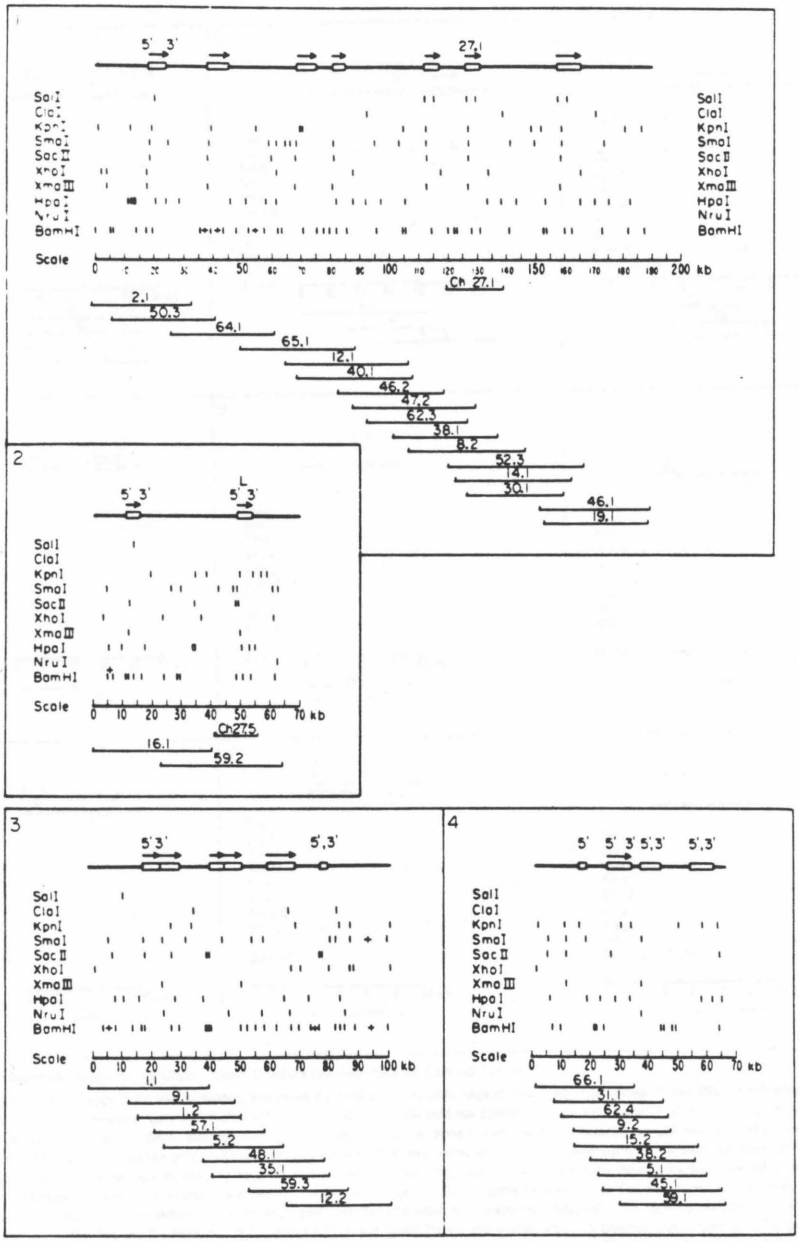
The Class I Genes in Certain Subgroups May Be Duplicated in Some Inbred Mouse Strains and Deleted from Others

Serologic studies indicate that the L gene is expressed in some inbred mouse strains (BALB/c [L^d], SWR [L^q]), but not in others (B6-Ly1^a [H-2^b], B6-H-2^k [H-2^k], ASW [H-2^s]) (Hansen et al., 1979). These and related observations raise the possibility that the L gene is present in BALB/c mice (H-2^d haplotype) and in SWR mice (H-2^q haplotype) but has been deleted from the genomes of B6-Ly1^a (H-2^b haplotype), B6-H-2^k (H-2^k haplotype) and ASW (H-2^s haplotype) mice. To test this possibility, we have isolated a low-copy-number probe from the 5′ flanking sequence of the L gene in the phage clone 27.5 (Fig. 3). We have used this probe to examine the DNA of mice of several haplotypes by Southern blot analysis (Figure 6). When DNA from mice of the H-2^d and H-2^q haplotypes is digested with Bam HI, a 19 kb band is observed. The size of this band corresponds to the Bam HI fragment in the restriction map for the L gene expected to hybridize to the probe (cluster 2, Figure 3). This Bam HI fragment is not found in DNAs from mice of the H-2^b, H-2^k and H-2^s haplotypes (Figure 6). Analysis with a second restriction enzyme reveals a similar pattern. At least two Kpn fragments with cross-reactive sequences, 12 kb and 18 kb, are present in the H-2^d haplotype DNA. The 12 kb band corresponds to the L gene (Figure 3), which does not appear to be present in DNAs of the H-2^b and H-2^k haplotypes (Figure 6).

These data suggest that the BALB/c mouse has at least two L-like genes (and probably more since the 18 kb band gives a stronger hybridization signal to our L-gene probe than the identified 12 kb L gene fragment). Accordingly, the L gene corresponding to clone 27.5 is missing from or mutated in the haplotypes that fail to express the L antigen. The low-copy-number 5′ probe from the L-gene cluster appears to detect two or even three homologous restriction enzyme fragments. With certain enzymes such as Bam HI, these homologous regions appear to be identical in size, that is, 19 kb (see Table 1). With other enzymes such as Kpn, two or more restriction fragments are seen. Thus it appears that the L subgroup of class I genes has undergone gene duplications and deletions in closely related inbred strains of mice. The 18 kb Kpn fragment or fragments have not been identified in our 54 cosmid clones. We are rescreening our cosmid library with the 5′ flanking probe to isolate clones containing this fragment.

Similar findings have been made for an intergenic region from cluster 1. A low-copy-number probe has been isolated from the region between genes 2 and 3 that hybridizes to six restriction fragments of BALB/c

Cell
492


1
5' 3'
27.1
SalI
ClaI
KpnI
SmaI
SacII
XhoI
XmaIII
HpaI
NruI
BamHI
Scale
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 kb
Ch 27.1
2.1
50.3
64.1
65.1
12.1
40.1
46.2
47.2
62.3
38.1
8.2
52.3
14.1
30.1
46.1
19.1
2
5' 3'
L
5' 3'
SalI
ClaI
KpnI
SmaI
SacII
XhoI
XmaIII
HpaI
NruI
BamHI
Scale
0 10 20 30 40 50 60 70 kb
Ch 27.5
16.1
59.2
3
5' 3'
5', 3'
SalI
ClaI
KpnI
SmaI
SacII
XhoI
XmaIII
HpaI
NruI
BamHI
Scale
0 10 20 30 40 50 60 70 80 90 100 kb
1.1
9.1
1.2
57.1
5.2
48.1
35.1
59.3
12.2
4
5' 5' 3' 5',3' 5',3'
SalI
ClaI
KpnI
SmaI
SacII
XhoI
XmaIII
HpaI
NruI
BamHI
Scale
0 10 20 30 40 50 60 70 kb
66.1
31.1
62.4
9.2
15.2
38.2
5.1
45.1
59.1

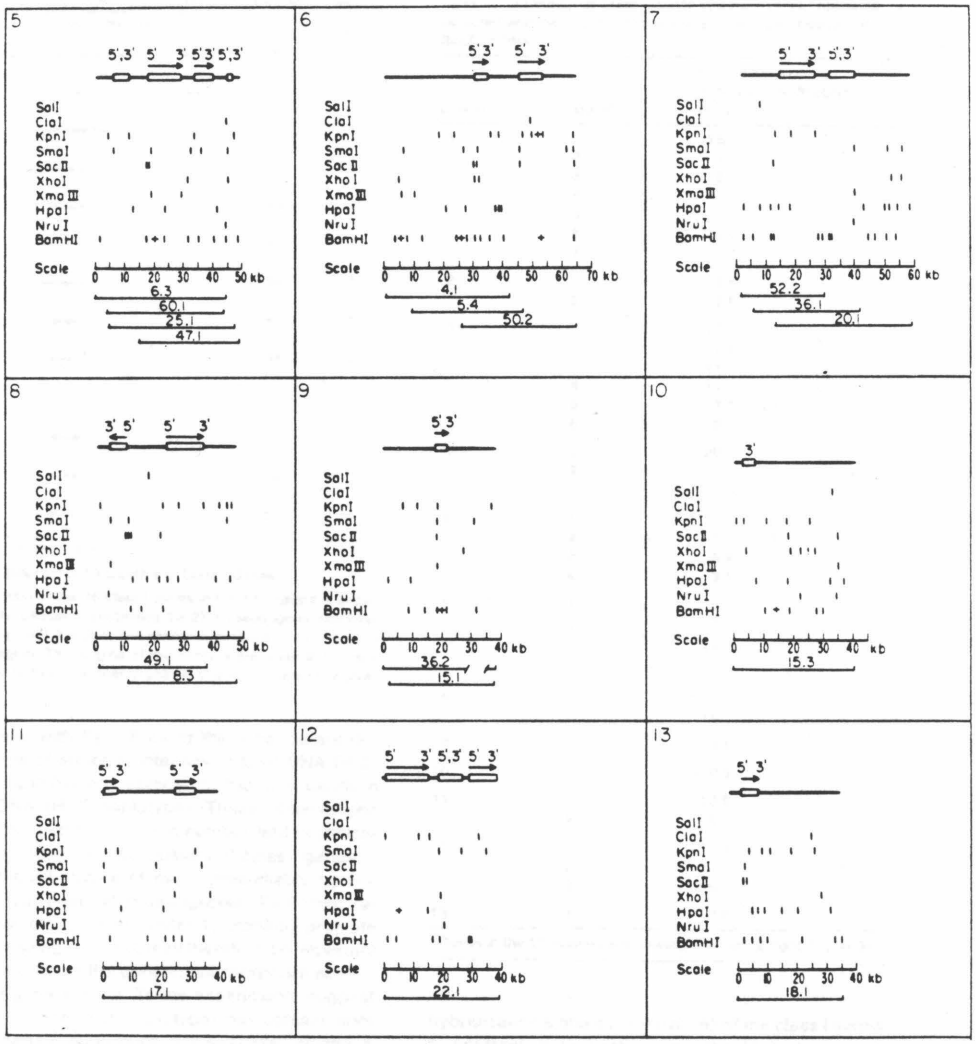


Figure 3. Restriction Maps for 13 Class I Gene Clusters Derived from 54 Cosmid Clones

Cosmid clones were mapped for the indicated enzymes by single and double digestions. Electrophoresis of the DNA fragments was on 20 cm long, horizontal 0.55% agarose gels in F buffer (40 mM Tris–acetate, 20 mM Na acetate, 1 mM EDTA [pH 7.2]) for 520 volt hr in parallel with a mixture of phage λ DNA and λ DNA restriction fragments (Sal I, Hind III, Sma I, Eco RI) as molecular weight markers. Gels were stained with ethidium bromide (1 μg/ml), destained in H_2O and photographed. DNA fragments were then transferred to nitrocellulose filters and hybridized with the 5′ or 3′ class I cDNA probes to identify restriction fragments containing class I coding regions (open boxes). Cosmid clones that showed overlapping restriction maps were aligned, and the composite maps for 13 class I gene clusters (1 to 13) are given. A + between two restriction sites indicates the presence of additional recognition sites for the enzymes that were not mapped. The regions contained in individual cosmid clones as well as in the two phage clones Ch27.1 and Ch27.5 are given below the maps, with the cosmid clone number. The gap in clone 36.2 (cluster 9) indicates the position where a deletion probably occurred during propagation of the clone. The 375 bp Eco RI–Bgl II fragment subcloned from phage 27.5 (see text) is indicated by a small box on the left side of the 27.5 map (cluster 2). The arrows above the coding regions indicates 5′ to 3′ directions. In cases where the 5′ to 3′ direction could not be determined, no arrow is given. With one exception, all of the genes that could be oriented show the same 5′ to 3′ direction in individual clusters.

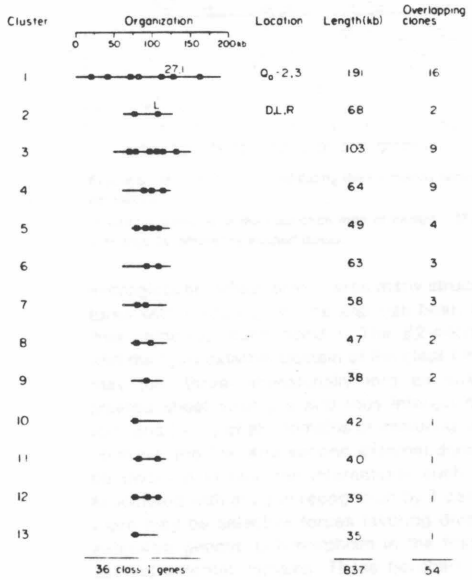


Figure 4. A Diagram of 13 Clusters of Class I Genes

The organization of the 36 class I genes in the 13 clusters is shown schematically. Cluster 1, containing the 27.1 pseudogene, has been located to the Qa-2,3 region, and cluster 2, containing the L gene, to the D-L-R region. The lengths of the clusters are given in kilobase pairs together with the number of cosmid clones contained in a given cluster.

DNA cleaved with Kpn. Four of these bands are deleted, one is of stronger intensity in C3H DNA (H-2^k haplotype) and five are replaced by four new bands in B6-Ly1^a DNA (H-2^b haplotype). These observations suggest that changes in gene number and organization occur throughout the clusters of class I genes in different inbred strains of mice, presumably by homologous but unequal crossing-over. Extensive regions of homology, as in cluster 1, probably promote unequal crossing-over and may therefore be important in the generation of the genetic polymorphism seen in the class I gene.system. These experiments suggest that BALB/c mice (H-2^d haplotype) may contain more genes in certain subgroups (for example, cluster 1 and the L subgroup) than are found in the genomes of mice of the H-2^k and H-2^b haplotypes.

The Exon Encoding the Third External Domain Is More Highly Conserved than Exons Encoding the First and Second External Domains

Southern blot analyses of our class I cosmid clones with the 5′ and 3′ cDNA probes (Figure 2) suggest that the exon encoding the third external domain is far more highly conserved than the exons encoding the first and second external domains. Hybridization with the 3′ probe appears to give equally strong hybridization signals to virtually all of the class I genes. In contrast, hybridization with the 5′ probe reveals intensity differences of two orders of magnitude among individual class I genes (Figure 7). These observations are consistent with sequence data derived from protein and DNA analyses which also suggest that the exon encoding the third external domain is more highly conserved than those encoding the first and second external domains. This finding has several functional implications. Since the exon encoding the third external domain shows significant sequence homology to immunoglobulin constant region domains, it is attractive to argue that the third external domain is conserved in order to interact efficiently with β2-

Table 1. Correlation of Class I Genes and the Bam HI Fragments Detected with the 3′ cDNA Probe in a Southern Blot Analysis of BALB/c DNA

Cluster	Gene[a]	Bam HI Fragment Hybridizing to 3′ cDNA Probe (kb)
1	1	15.5
	2	1
	3	5.7
	4	3.4
	5	5.7
	6	2.7
	7	2.7
2	1	2.6
	2	2.6
3	1	10.5
	2	2.7
	3	11
	4	11
	5	3.3 + 3.1
	6	6
4	1	20
	2	20
	3	15
5	1	15.5
	2	7.8 + 1.1
	3	5.2
	4	3.1
6	1	3
	2	3
7	1	15
	2	12.5
8	1	>11
	2	15
9	1	3.1
10	1	>10.5
11	1	12.5
	2	9
12	1	12.5
	2	10
	3	>8.5
13	1	2.6

[a] Genes in the 13 clusters are counted from left to right (Figure 3).

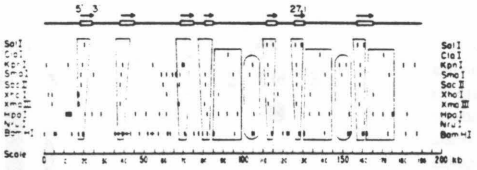


Figure 5. Map of Cluster 1 Indicating the Homology between Individual Genes

Repetitive patterns in the restriction map of cluster 1 (Figure 3) are indicated by differently shaped boxes.

microglobulin, which also shares many structural features with immunoglobulins and has been termed a free immunoglobulin domain. The β2-microglobulin and the third external domain of the class I molecules may fold three-dimensionally into an antibody β-pleated sheet structure and thus interact much like light and heavy chain domains of immunoglobulins. In contrast, the first and second external domains may be involved in cell–cell interactions such as those associated with antigen recognition by T cells. Hence there may be selective forces favoring diversity and extensive genetic polymorphism in the first and the second external domains. These Southern blot data confirm and extend the homology relationships suggested by limited amino acid studies (Nathenson et al., 1981) and DNA sequence studies (Steinmetz et al., 1981a and 1981b) of class I polypeptides and genes.

Class I Genes and the Major Histocompatibility Complex

We have identified 36 distinct class I genes in BALB/c mice, which encompass 837 kb of DNA. We have pooled DNAs from cosmid clones encoding the 36 class I genes, digested them with the restriction enzyme Bam HI and compared by Southern blot analyses these pooled DNAs to similarly treated DNAs from BALB/c liver and sperm. Most of the bands in genomic DNA that hybridize with a 3′ class I cDNA probe also are found in the pooled DNA from the cosmid clones. Thus we believe these clones contain most of the mouse class I genes cross-hybridizing to the class I cDNA probes. There is a possibility that additional class I genes will not be detected by our class I cDNA probes. However, our 3′ cDNA probe readily detects the Tla gene (R. Goodenow, unpublished observation) that encodes a T-cell differentiation antigen presumably distantly related to the classical transplantation antigens. These observations suggest that most class I genes will be detected by our 3′ cDNA probe.

An intriguing question is what fraction of the 36 class I genes are pseudogenes? Of the first two completely sequenced class I genes, one was a pseudogene (27.1), and the other was an L^d gene (27.5). We should be able to answer this question with the DNA

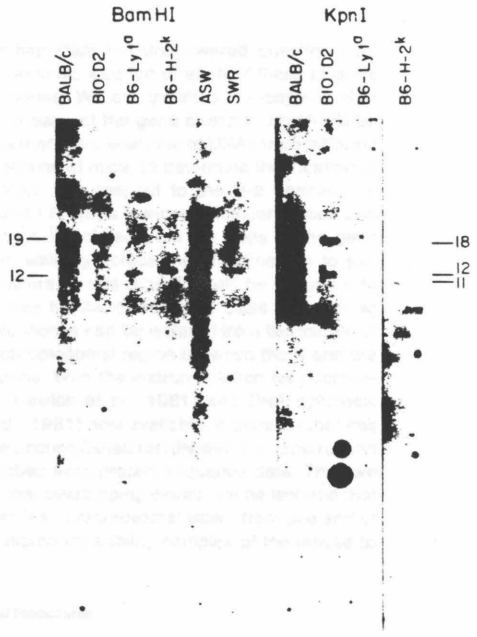


Figure 6. A Southern Blot Analysis of DNA from the d, b, k, s, and q Haplotypes with a Low-Copy-Number Probe Located 5′ to the L^d Gene

A low-copy-number DNA fragment was identified at the end of the insert in the phage clone Ch27.5 containing the L^d gene by virtue of the fact that it did not hybridize when total BALB/c mouse liver DNA was used as a probe (Steinmetz et al., 1980). This 375 bp Eco RI–Bgl II fragment, indicated in the map of Ch27.5 in Figure 3 as a small box, was subcloned into M13mp701; single-stranded phage DNA was isolated, and short DNA fragments complementary to the M13 insert were synthesized with α-^{32}P-labeled dATP, dGTP, dCTP and dTTP (>400 Ci/mmole) under the conditions described previously (Steinmetz et al., 1981b). The labeled single-stranded DNA fragments were separated from the template DNA on a Bio-Gel A150m column in 10 mM NaOH and 1 mM EDTA, recovered by ethanol precipitation and used as a hybridization probe. Hybridization conditions were as described in Experimental Procedures, and two stringent washes were performed after hybridization in 1 × SSC and 0.1% SDS at 68°C for 15 min each. Exposure was with an intensifying screen for 5 days. The nitrocellulose filters contained Bam HI– or Kpn I–digested mouse liver DNA (5–10 μg each) from BALB/c (d haplotype), B10.D2 ($H\text{-}2^o$ alleles crossed into a b haplotype mouse), B6-$Ly1^a$ (b haplotype), B6-$H\text{-}2^k$ ($H\text{-}2^k$ alleles crossed into a b haplotype mouse), ASW (s haplotype) and SWR (q haplotype) mice. The sizes of the hybridizing fragments are given in kilobase pairs.

sequence and gene transfer studies on the class I genes that are under way. The existence of multiple, closely related class I genes also raises the possibility that there may be examples in addition to Qa and Tla of tissue-specific expression of individual class I genes (see Rouse and Weissman, 1981).

Simple but admittedly approximate calculations demonstrate that we have probably cloned a signif-

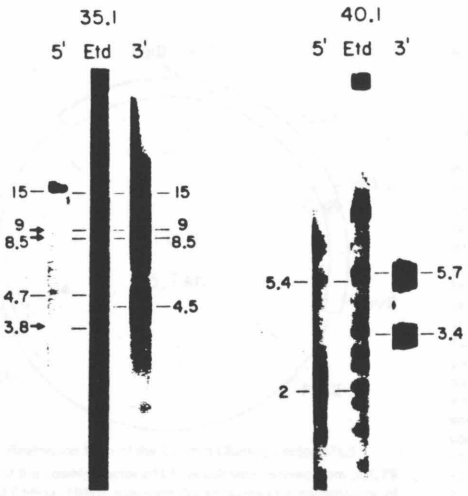


Figure 7. A Southern Blot Analysis of Two Cosmid Clones with the 5′ and 3′ Class I cDNA Probes

Cosmid clones 35.1 and 40.1 (0.25 μg each) were cleaved with a mixture of Sma I plus Cla I and with Bam HI, respectively, separated on a 0.55% agarose gel in 40 mM Tris–HCl, 20 mM sodium acetate and 1 mM EDTA (pH 7.2), transferred to nitrocellulose filters and hybridized with the 5′ and the 3′ class I cDNA probes (both probes had a specific activity of about 2×10^7 cpm/μg). Hybridization was as described in Experimental Procedures, and the filters were washed twice in 3× SSC and 0.1% SDS, and twice in 0.1× SSC and 0.1% SDS at room temperature for 20 min each. Exposure was overnight without a screen. Lane Etd, restriction fragments obtained from the digested clones after staining with ethidium bromide; lane 5′, fragments identified by hybridization with the 5′ probe; lane 3′, with the 3′ probe. The sizes of the fragments that hybridize are given in kilobase pairs. Arrows indicate fragments that showed weak hybridization on the original autoradiogram. The 5.7 kb and the 5.4 kb fragments derived from clone 40.1 cover gene 3 in cluster 1 (genes counted from left to right; the 5.4 kb fragment extends into the cosmid vector) and the fragments 3.4 kb and 2 kb in length from clone 40.1 cover gene 4. For clone 35.1 (cluster 3), the following correlations exist: 4.5 kb fragment (extends into the vector), gene 3; 9 kb fragment, gene 4; 8.5 and 3.8 kb fragments, gene 5; 15 kb fragment, gene 6. The weakly hybridizing 4.7 kb fragment (mostly vector DNA) appears because of contaminating pBR322 sequences in the 5′ cDNA probe.

icant fraction of the mouse major histocompatibility complex. The mouse genome encompasses 1600 centiMorgans (cM) of DNA (V. McKusick, personal communication), which is equivalent to 3×10^6 kb of DNA in the haploid mouse genome. If crossing-over occurred randomly throughout the genome, then 1 cM would equal about 2000 kb of DNA. Thus the H-2 complex (0.3 cM) may span about 600 kb of DNA, and the Tla complex (1 cM) may include about 2000 kb of DNA (see Figure 1). Thus it is likely that we have isolated a substantial fraction of DNA from the major histocompatibility complex, even if the Tla complex is included, as it should be, within the major histocompatibility complex.

Another important but unanswered question concerns the genomic location of each of these clusters of class I genes. We are isolating low-copy-number probes from each of the gene clusters, which will be used in Southern blot analyses of DNAs from different congenic strains of mice, to determine the location of these clusters with respect to the H-2 complex on chromosome 17. These low-copy-number probes also will be useful, if isolated from the ends of the gene clusters, in walking across the chromosome to join various clusters. In this regard, it will be important to isolate probes for the class II and class III genes so that cosmid clones can be isolated from the middle of the large chromosomal region between the K and the D-L-R regions. With the instrumentation for microsequencing (Hewick et al., 1981) and DNA synthesis (Hood et al., 1981) now available, it appears that this task will be accomplished rapidly with the construction of DNA probes from protein sequence data. Thus we anticipate that overlapping clones will be isolated that should permit a "chromosomal walk" from one end of the major histocompatibility complex of the mouse to the other.

Experimental Procedures

Materials

Restriction enzymes were purchased from New England BioLabs, Bethesda Research Laboratories and Boehringer Mannheim. Bacterial alkaline phosphatase was obtained from Bethesda Research Laboratories, and T4 DNA ligase from M. C. Alonso; α-^{32}P-dNTPs (10 mCi/ml, >400 Ci/mmole) were obtained from Amersham or New England Nuclear. Nitrocellulose filters used for plating and screening of the cosmid library were purchased from Millipore (HA, 0.45 μm pore size, 125 mm in diameter), and those used for Southern blotting were from Schleicher and Schüll (BA85, 0.45 μm pore size). The cosmid vector pTL5 and the E. coli strain ED8767 (recA$^-$) were obtained from F. G. Grosveld and R. A. Flavell, M13mp701 from D. R. Bentley and E. coli 490A ($r_k^-m_k^-$met$^-$thr$^-$leu$^-$recA$^-$) from G. Hobom.

Construction of the Cosmid Library

The cosmid library was constructed from partially digested BALB/c sperm DNA and the cosmid vector pTL5 (Lund et al., 1982) according to the techniques described by Grosveld et al. (1981) and Ish-Horowicz and Burke (1981). The strategy to clone Mbo I–generated eucaryotic DNA fragments into the single Bgl II site of pTL5 (Figure 8) was outlined to us by F. G. Grosveld. Equal amounts of pTL5 were digested separately with Pvu II and Bst EII and were phenol-extracted. The two digests were combined, dialyzed against 10 mM Tris and 1 mM EDTA (pH 7.5) to remove inorganic phosphate that otherwise inhibits the bacterial alkaline phosphatase (BAP) and then treated with BAP. The BAP treatment was followed by a Bgl II digestion, and the two vector arms containing the cos sequence—namely, the 1.6 kb Bgl II–Pvu II and the 4.7 kb Bst EII–Bgl I fragment—were isolated by gel electrophoresis followed by purification by BD cellulose chromatography (Steinmetz et al., 1979). A sixfold molar excess of equimolar amounts of the vector arms was then ligated with partial Mbo I fragments of BALB/c sperm DNA, 30 to 50 kb in length, which had been isolated by size selection on a sucrose gradient (Maniatis et al., 1978). The ligation with T4 DNA ligase at 4000 U/ml was in 20 mM Tris–HCl (pH 7.5), 10 mM $MgCl_2$, 10 mM dithiothreitol, 1 mM spermidine–HCl and 1 mM ATP, at a DNA concentration of 250 μg/ml for 15 hr at 15°C. In vitro packaging was carried out by the procedures of Hohn and Murray (1977), with 5 λ of the ligation

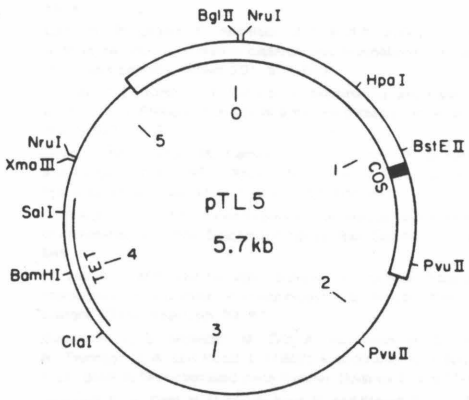


Figure 8. Restriction Map of the Cosmid Cloning Vector pTL5

The map of the cosmid vector pTL5, which was derived from pHC79 (Hohn and Collins, 1980), indicates the sites used for construction of the cosmid library (see Experimental Procedures) and those for the enzymes used in the mapping of the cosmid clones. pTL5 has no recognition sites for Sma I, Sac II, Kpn I and Xho I. The open box indicates the phage λ fragment containing the cos sequence, and TET gives the location of the tetracycline resistance gene in the pBR322 part of the cosmid vector. The scale is given in kilobase pairs.

mixture per 50 μl packaging extract prepared from the strains of Sternberg et al. (1977). Three milliliters of SMC (Hohn and Murray, 1977) were then added per packaging reaction, followed by 6 ml of plating cells (E. coli 490A or ED8767). Transduction was performed for 20 min at room temperature, followed by the addition of 60 ml of L broth, and tetracycline resistance was induced by shaking at 37°C for 45 min. The cells were pelleted and resuspended in 8 ml L broth, and 600 λ aliquots were plated on Millipore nitrocellulose filters (125 mm in diameter) on fresh F plates (Hanahan and Meselson, 1980) containing 5 μg/ml tetracycline. After 3 hr of incubation at 37°C, the filters were transferred to F plates containing 10 μg/ml tetracycline and incubated overnight at 30°C. The initial incubation on plates containing only 5 μg/ml instead of 10 μg/ml tetracycline increased the yield of resistant colonies by about a factor of 3. Filters were then replica plated and stored at −70°C as described by Hanahan and Meselson (1980).

With 4 μg of size-selected BALB/c sperm DNA fragments, we constructed a cosmid library containing about 250,000 colonies distributed on 65 nitrocellulose filters. No difference was observed in the use of the E. coli 490A or ED8767 host cells.

Screening of the Cosmid Library

For screening the library, we isolated the 349 bp Pst I–Pst I fragment from pH-2I, the 442 bp Hha I–Sac I fragment (3′ class I cDNA probe, Figure 2) from pH-2IIa and the 287 bp Pst I–Pst I fragment (5′ class I cDNA probe, Figure 2) from pH-2III (Steinmetz et al., 1981a) by agarose gel electrophoresis and BD cellulose purification. These fragments were ligated to themselves, then labeled by nick translation (Maniatis et al., 1975; Rigby et al., 1977) to a specific activity of about 4×10^7 cpm/μg. Duplicates of the 65 filters containing the cosmid library were prepared for hybridization essentially as described by Grosveld et al. (1981) and hybridized in 150 ml of 3× SSC, 10× Denhardt's solution (Denhardt, 1966), 50 μg/ml denatured salmon sperm DNA, 200 μg denatured pBR322–Hae III fragments, 10 μg/ml each of poly(A), poly(G) and poly(C) and 0.1% SDS with 2 μg of a mixture of the three cDNA probes for 18 hr at 68°C in a shaking water bath. The filters were then washed six times for 15 min each with 300 ml of 3× SSC, 10× Denhardt's solution and 0.5% SDS, and three times for 15 min each with 300 ml of 1× SSC and 0.1% SDS. Filters were exposed overnight with an intensifying screen; colonies that were positive on both filters of the duplicate set were picked and rescreened as described by Grosveld et al. (1981), with use of the supernatant from the first round of hybridization.

Preparation of Plasmid DNA from the Cosmid Clones

Plasmid DNA from the cosmid clones was isolated by the cleared lysate method described by Davis et al. (1980) with some modifications. This method gave similar or better yields than the alkaline extraction method (Birnboim and Doly, 1979). We infected 250 ml of L broth containing 10 μg/ml of tetracycline in a 2 liter flask with an exponentially growing culture and incubated the solution overnight in an air shaker with vigorous agitation. Cells were pelleted (25 min at 4300 rpm in a Beckman J6 centrifuge), washed once with 35 ml 10 mM Tris–HCl (pH 7.5) and 1 mM EDTA (TE buffer) and resuspended in 5 ml of sucrose solution by tossing up and down with a pipette. Then 1.5 ml of lysozyme solution was added, the suspension was incubated for 10 min on ice and lysis was achieved by the addition of 4 ml of Triton X-100 solution, a brief treatment on the vortex mixer and incubation for 15 min at 37°C (see Davis et al., 1980, for the composition of these solutions). Cleared lysates were obtained by a spin at 15,000 rpm for 40 min in the Sorvall SS34 rotor. Then 1.059 g of cesium chloride and 0.059 ml of ethidium bromide (10 mg/ml) were added per milliliter of supernatant, and supercoiled cosmid DNA was isolated by two equilibrium centrifugations in the Beckman VTi50 rotor (20 hr at 45,000 rpm). Ethidium bromide was extracted five times with a 1:1 solution of isopropanol/n-butanol (equilibrated with a saturated cesium chloride solution in TE buffer), and the DNA was dialyzed against TE buffer. About 250–500 μg of recombinant cosmid DNA were obtained from a 250 ml culture.

Southern Blot Hybridization

DNA blots were prepared (Southern, 1975) and hybridized for 16 hr at 68°C in a rotisserie oven in 5× SSC, 10× Denhardt's solution, 100 μg/ml denatured salmon sperm DNA, 10 μg/ml each of poly(A), poly(G) and poly(C), 5 mM EDTA and 0.1% SDS with nick-translated probes (at 5–20 ng/ml) or single-stranded probes (at 2×10^6 cpm/ml) labeled by second strand synthesis in M13. Filters were washed as indicated in the figure legends.

Acknowledgments

This work was supported by a grant from the National Institutes of Health. We thank Henry Sun for help in the preparation of plasmid DNA from the cosmid clones; Tim Hunkapiller, Mary Macchi and Dr. Jeff Frelinger for some of the mouse DNAs used in Southern blot analyses; Beverly Taylor Sher for the BALB/c sperm DNA; Doug Fisher for the preparation of the M13mp701 DNA used for subcloning, and Bernita Larsh for the preparation of the manuscript. We also thank Dr. David Sachs for informative discussions; Drs. Roger Perlmutter, Ellen Kraig and Johanna Griffin, and Steve Crews and Mitch Kronenberg for comments on the manuscript, and Dr. Bertrand Jordan for the communication of unpublished results. M. S. is the recipient of a fellowship from the Deutsche Forschungsgemeinschaft.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received December 9, 1981; revised December 31, 1981

References

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

Brégégère, F., Abastado, J. P., Kvist, S., Rask, L., Lalanne, J. L., Garoff, H., Cami, B., Wiman, K., Larhammar, D., Peterson, P. A.,

Gachelin, G., Kourilsky, P. and Dobberstein, B. (1981). Structure of C-terminal half of two H-2 antigens from cloned mRNA. Nature *292*, 78–81.

Cami, B., Brégégère, F., Abastado, J. P. and Kourilsky, P. (1981). Multiple sequences related to classical histocompatibility antigens in the mouse genome. Nature *291*, 673–675.

Coligan, J. E., Kindt, T. J., Uehara, H., Martinko, J. and Nathenson, S. G. (1981). Primary structure of a murine transplantation antigen. Nature *291*, 35–39.

Davis, R. W., Thomas, M., Cameron, J., St. John, T. P., Scherer, S. and Padgett, R. A. (1980). Rapid DNA isolations for enzymatic and hybridization analysis. Meth. Enzymol. *65*, 404–411.

Denhardt, D. T. (1966). A membrane-filter technique for the detection of complementary DNA. Biochem. Biophys. Res. Commun. *23*, 641–646.

Flaherty, L. (1980). The Tla region antigens. In The Role of the Major Histocompatibility Complex in Immunology, M. E. Dorf, ed. (New York: Garland STPM Press), pp. 33–57.

Goodenow, R. S., McMillan, M., Örn, A., Nicholson, M., Davidson, N., Frelinger, J. A. and Hood, L. (1982). Identification of a BALB/c H-2L^{o} gene by DNA-mediated gene transfer. Science *215*, 677–679.

Grosveld, F. G., Dahl, H. H. M., de Boer, E. and Flavell, R. A. (1981). Isolation of β-globin-related genes from a human cosmid library. Gene *13*, 227–237.

Hanahan, D. and Meselson, M. (1980). Plasmid screening at high colony density. Gene *10*, 63–67.

Hansen, T. H., Ivanyi, P., Levy, R. B. and Sachs, D. H. (1979). Cross-reactivity among the products of three nonallelic H-2 loci, H-2L^{d}, H-2D^{o}, and H-2K^{k}. Transplantation *28*, 339–342.

Hansen, T. H., Ozato, K., Melino, M. R., Coligan, J. E., Kindt, T. J., Jandinski, J. J. and Sachs, D. H. (1981). A gene cluster in the H-2D region: evidence in two haplotypes for at least three D region-encoded molecules, D, L, and R. J. Immunol. *126*, 1713–1716.

Hewick, R. M., Hunkapiller, M. W., Hood, L. E. and Dreyer, W. J. (1981). A gas-liquid solid phase peptide and protein sequenator. J. Biol. Chem. *256*, 7990–7997.

Hohn, B. and Collins, J. (1980). A small cosmid for efficient cloning of large DNA fragments. Gene *11*, 291–298.

Hohn, B. and Murray, K. (1977). Packaging recombinant DNA molecules into bacteriophage particles in vitro. Proc. Nat. Acad. Sci. USA *74*, 3259–3263.

Hood, L., Hunkapiller, M., Hewick, R., Giffin, C. E. and Dreyer, W. J. (1981). Microchemical instrumentation. J. Supramol. Structure Cell. Biochem. *17*, 27–36.

Ish-Horowicz, D. and Burke, J. F. (1981). Rapid and efficient cosmid cloning. Nucl. Acids Res. *9*, 2989–2998.

Klein, J. (1975). Biology of the Mouse Histocompatibility Complex (Berlin: Springer Verlag).

Kvist, S., Brégégère, F., Rask, L., Cami, B., Garoff, H., Daniel, F., Wiman, K., Larhammar, D., Abastado, J. P., Gachelin, G., Peterson, P. A., Dobberstein, B. and Kourilsky, P. (1981). cDNA clone coding for part of a mouse H-2^{d} major histocompatibility antigen. Proc. Nat. Acad. Sci. USA *78*, 2772–2776.

Leder, A., Swan, D., Ruddle, F., D'Eustachio, P. and Leder, P. (1981). Dispersion of α-like globin genes of the mouse to three different chromosomes. Nature *293*, 196–200.

Lopez de Castro, J. A., Orr, H. T., Robb, R. J., Kostyk, T. G., Mann, D. L. and Strominger, J. L. (1979). Complete amino acid sequence of a papain-solubilized human histocompatibility antigen, HLA-B7. I. Isolation and amino acid composition of fragments and of tryptic and chymotryptic peptides. Biochemistry *18*, 5704–5711.

Lund, T., Grosveld, F. G. and Flavell, R. A. (1982). Isolation of transforming DNA by cosmid rescue. Proc. Nat. Acad. Sci. USA *79*, 520–524.

Malissen, M., Malissen, B. and Jordan, B. R. (1982). Exon–intron organization and complete nucleotide sequence of an HLA gene. Proc. Nat. Acad. Sci. USA, in press.

Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K. and Efstratiadis, A. (1978). The isolation of structural genes from libraries of eucaryotic DNA. Cell *15*, 687–701.

Maniatis, T., Jeffrey, A. and Kleid, D. G. (1975). Nucleotide sequence of the rightward operator of phage λ. Proc. Nat. Acad. Sci. USA *72*, 1184–1188.

Michaelson, J., Flaherty, L., Vitetta, E. and Poulik, M. D. (1977). Molecular similarities between the Qa-2 alloantigen and other gene products of the 17th chromosome of the mouse. J. Exp. Med. *145*, 1066–1070.

Moore, K. W., Sher, B. T., Sun, Y. H., Eakle, K. and Hood, L. (1982). DNA sequence of a gene encoding a BALB/c mouse L^{d} transplantation antigen. Science *215*, 679–682.

Nathenson, S. G., Uehara, H., Ewenstein, B. M., Kindt, T. J. and Coligan, J. E. (1981). Primary structural analysis of transplantation antigens of the murine H-2 major histocompatibility complex. Ann. Rev. Biochem. *50*, 1025–1052.

Ploegh, H. L., Orr, H. T. and Strominger, J. L. (1980). Molecular cloning of a human histocompatibility antigen cDNA fragment. Proc. Nat. Acad. Sci. USA *77*, 6081–6085.

Ploegh, H. L., Orr, H. T. and Strominger, J. L. (1981). Major histocompatibility antigens: the human (HLA-A, -B, -C) and murine (H-2K, H-2D) class I molecules. Cell *24*, 287–299.

Rigby, P. W. J., Dieckmann, M., Rhodes, C. and Berg, P. (1977). Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J. Mol. Biol. *113*, 237–251.

Rouse, R. V. and Weissman, I. L. (1981). Microanatomy of the thymus: its relationship to T cell differentiation. Ciba Foundation Symposium 84 (London: Pitman Medical), in press.

Soloski, M. J., Uhr, J. W., Flaherty, L. and Vitetta, E. S. (1981). Qa-2, H-2K, and H-2D alloantigens evolved from a common ancestral gene. J. Exp. Med. *153*, 1080–1093.

Sood, A. K., Pereira, D. and Weissman, S. M. (1981). Isolation and partial nucleotide sequence of a cDNA clone for human histocompatibility antigen HLA-B by use of an oligodeoxynucleotide primer. Proc. Nat. Acad. Sci. USA *78*, 616–620.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. *98*, 503–517.

Stanton, T. H. and Hood, L. (1980). Biochemical identification of the Qa-1 alloantigen. Immunogenetics *11*, 309–314.

Steinmetz, M., Frelinger, J. G., Fisher, D., Hunkapiller, T., Pereira, D., Weissman, S. M., Uehara, H., Nathenson, S. and Hood, L. (1981a). Three cDNA clones encoding mouse transplantation antigens: homology to immunoglobulin genes. Cell *24*, 125–134.

Steinmetz, M., Höchtl, J., Schnell, H., Gebhard, W. and Zachau, H. G. (1980). Cloning of V region fragments from mouse liver DNA and localization of repetitive DNA sequences in the vicinity of immunoglobulin gene segments. Nucl. Acids Res. *8*, 1721–1729.

Steinmetz, M., Moore, K. W., Frelinger, J. G., Sher, B. T., Shen, F.-W., Boyse, E. A. and Hood, L. (1981b). A pseudogene homologous to mouse transplantation antigens: transplantation antigens are encoded by eight exons that correlate with protein domains. Cell *25*, 683–692.

Steinmetz, M., Zachau, H. G. and Mach, B. (1979). Cloning of immunoglobulin kappa light chain genes from mouse liver and myeloma MOPC173. Nucl. Acids Res. *6*, 3213–3229.

Sternberg, N., Tiemeier, D. and Enquist, L. (1977). In vitro packaging of a λ Dam vector containing Eco RI fragments of E. coli and phage P1. Gene *1*, 255–280.

Zinkernagel, R. M. and Doherty, P. C. (1980). MHC-restricted cytotoxic T cells: studies on the biological role of polymorphic major transplantation antigens determining T-cell restriction-specificity, function, and responsiveness. Adv. Immunol. *27*, 51–177.

Chapter Three

GENETIC MAPPING IN THE MAJOR HISTOCOMPATIBILITY COMPLEX BY RESTRICTION ENZYME SITE POLYMORPHISM: MOST CLASS I GENES MAP TO THE TLA COMPLEX

Published in

Proceedings of the National Academy of Sciences USA **80,** 3425-3429, 1983

Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 3425–3429, June 1983
Immunology

Genetic mapping in the major histocompatibility complex by restriction enzyme site polymorphisms: Most mouse class I genes map to the *Tla* complex

(Southern blotting)

AŞTAR WINOTO, MICHAEL STEINMETZ, AND LEROY HOOD

Division of Biology, California Institute of Technology, Pasadena, California 91125

Contributed by Leroy Hood, February 15, 1983

ABSTRACT **From a genomic library constructed from sperm DNA of the inbred BALB/c mouse, we previously isolated 54 cosmid clones that contain 36 class I genes and can be divided by restriction map analyses into 13 gene clusters. We have isolated single- and low-copy DNA probes from each of these clusters to visualize restriction enzyme site polymorphisms in the DNAs from various congeneic and recombinant congeneic mice. These polymorphisms permit us to map each of the 13 cosmid clusters to a precise location in the major histocompatibility complex of the mouse. Thirty-one of 36 class I genes map into the *Tla* complex of the major histocompatibility complex whereas the remaining 5 genes map to the *H-2* complex. Thus, all 36 class I genes are located in the major histocompatibility complex. Analysis of the number of restriction enzyme fragments visualized by the single- and low-copy DNA probes suggests that the class I genes in different inbred strains of mice probably undergo gene duplications and deletions, presumably by homologous but unequal crossing-over.**

The major histocompatibility complex (MHC) of the mouse contains at least three gene families, denoted class I, class II, and class III (1–4). The molecules encoded by the class I genes fall into two categories by virtue of difference in cellular distribution, the extent of their serological polymorphisms, and their functions. The class I genes designated *K*, *D*, and *L* encode transplantation antigens that are found on the cell surface of most nucleated cells and are highly polymorphic (5). These integral membrane proteins serve as restricting elements that permit cytotoxic T cells to recognize viral or tumor antigens on the surface of infected or transformed cells (6). The class I genes denoted *Qa-2,3*, *Tla*, and *Qa-1* encode antigens that are present on certain hematopoietic cells. They are far less polymorphic than their transplantation antigen counterparts and their functions are unknown (7).

The availability of inbred strains of mice as well as congeneic mice, which are genetically identical except for the genes of the MHC, has permitted immunogeneticists to use serologic polymorphisms and recombinational analyses to identify at least six class I genes that map to two distinct complexes of MHC. The *H-2* complex comprises the proximal part of the MHC on chromosome 17 and includes the *K* gene as its left-hand boundary and the *D* and *L* genes as its right-hand boundary. The class II and III gene families lie between the *K* and *D* regions of the *H-2* complex. The *Tla* complex comprises the distal part of the MHC and includes the *Qa-2,3*, *Tla*, and *Qa-1* genes. Various inbred strains of mice are distinguished by having different alleles at the six serologically defined loci and these constellations of alleles are denoted haplotypes. The MHC encompasses about 2 centimorgans of DNA and this may be equivalent to as much as 2,000–4,000 kilobases of DNA (1, 2, 8).

We have recently isolated 54 genomic clones from a cosmid library constructed from BALB/c sperm DNA. These 54 cosmid clones contain 36 distinct class I genes, which map into 13 gene clusters (9). In this paper, we report the use of the technique of genetic mapping by restriction enzyme site polymorphisms to locate the positions of all 13 of the class I gene clusters within the MHC. These results are in complete agreement with those provided by the serological identification of class I genes by gene transfer that identified six (*K*, *D*, *L*, *Qa-2,3*, and two *TL*) of the 36 class I genes (10). Our results show that 31 of the 36 class I genes are located in the *Tla* complex.

MATERIALS AND METHODS

Materials. The sources of the restriction enzymes, T4 DNA ligase, [α-^{32}P]dNTPs, and nitrocellulose filters have been described (9).

Methods. Isolation of mouse DNA, plasmid DNA, and the procedure used for subcloning have been described (9, 11–13). Southern blot hybridization was carried out as follows. DNA blots were prepared and hybridized for 16 hr at 68°C in a rotisserie oven in 0.30 M NaCl/0.03 M Na citrate, pH 7.0/0.10% polyvinylpyrrolidone/0.10% Ficoll/0.1% bovine serum albumin/10% dextran sulfate/denatured salmon sperm DNA (100 μg/ml)/denatured BALB/c mouse DNA (100 μg/ml)/5 mM EDTA/0.1% $NaDodSO_4$ with nick-translated probes at 1×10^6 cpm/ml. Filters were washed twice with 0.30 M NaCl/0.03 M Na citrate, pH 7.0/0.1% $NaDodSO_4$ and twice with 15 mM NaCl/1.5 mM Na citrate, pH 7.0/0.1% $NaDodSO_4$ at 68°C for 20 min each. Dextran sulfate and denatured mouse DNA were omitted when nick-translated mouse DNA was used as a probe.

Two procedures have been used for the isolation of single- or low-copy probes. First, labeled mouse DNA was hybridized to electrophoresed restriction fragments of individual cosmid clones that were blotted onto nitrocellulose under conditions in which hybridization occurs only to fragments containing repetitive sequence elements (14). Those fragments failing to hybridize were then subcloned into phage M13 mp8 or plasmid pBR322. A second approach was used for those cases in which we were unable to identify single- or low-copy restriction fragments by this procedure. DNA fragments from cosmid clones were cleaved with frequent-cutting restriction enzymes such as *Sau*3A and the resulting small fragments were cloned into phage M13 mp8. Phage clones were then analyzed by a plaque hybridization procedure with total mouse DNA and the cosmid fragment as probes.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: MHC, major histocompatibility complex; kb, kilobase(s).

Table 1. Comparison between MHC alleles and polymorphic restriction fragments used for mapping in congeneic and

MHC allele*										
K	*D*	*Qa-2,3*	*Tla*	Strain	1 (*Kpn* I)	2 (*Bam*HI)	3 (*Eco*RI)	4 (*Kpn* I)	5 (*Kpn* I)	6 (*Kpn* I)
d	*d*	*a*	*c*	BALB/c	15.4	8	2.2	14	13	10.5
k	*k*	*b*	*b*	C3H	15.4		MB	14		16
k	*k*	*b*	*b*	AKR		10		14	13	
b	*b*	*a*	*b*	C57BL/6	MB	8	MB	14	10	13
b	*b*	*a*	*b*	C57BL/10						
k	*d*	*a*	*a*	A/WySn			2.2	—†	13	
s	*d*	*a*	*c*	A.TL						
k	*d*	*a*	*c*	A.AL		8				
q	*d*	*a*	*a*	A.QR						
b	*b*	*a*	*b*	A.BY			MB			
k	*d*	*a*	*a*	B10.A			2.2	—		
k	*b*	*a*	*b*	B10.AM		8	2.2			
k	*b*	*a*	*b*	B10.A(1R)				14		13
b	*d*	*a*	*a*	B10.A(3R)				—	13	
d	*b*		(*b*)	B10.BDR-1					10	
d	*d*	*a*	*c*	B10.D2						
q	*q*	*a*	*b*	B10.Q						
b	*b*	*a*	*a*	B6-Tla^a	MB	8	2.2	—	13	13
k	*k*		*b*	B6-H-2^k	15.4	10	MB		13	16
b	*b*	*b*	*b*	B6.K1	15.4	8	MB	14	13	16 and 13
b	*b*	*a*	*b*	B6.K2	MB	8	MB	14	13	13
Cluster locations					*Qa*	*D*	*Tla*	*Tla*	*Qa/Tla*	*D–Qa*

The probes used for mapping the clusters were cluster 1, 2.1-kb *Eco*RI/*Nru* I fragment of cosmid 65.1 subcloned into the *Eco*RI/*Nru* I sites of pBR322; cluster 2, *Sau*3A fragment of 8-kb *Bam*HI fragment from cosmid 59.2 subcloned into the *Bam*HI site of M13 mp8; cluster 3, 2.2-kb *Eco*RI fragment of cosmid 12.2 subcloned into the *Eco*RI site of M13 mp8; cluster 4, 2-kb *Kpn* I/*Sma* I fragment of cosmid 66.1 subcloned into the *Hinc*II site of M13 mp8; cluster 5, 2.6-kb *Xho* I/*Hpa* I fragment of cosmid 47.1; cluster 6, 2-kb *Sma* I fragment of cosmid 50.2 subcloned into the *Hinc*II site of M13 mp8; cluster 7, 3-kb *Xho* I fragment of cosmid 20.1 subcloned into the *Sma* I site of M13 mp8; cluster 8, 3-kb *Hpa* I fragment of cosmid 49.1 subcloned into the *Hinc*II site of M13 mp8; cluster 9, *Sau*3A fragment from 6.4-kb *Kpn*I fragment of cosmid 36.2 subcloned into the *Bam*HI

RESULTS AND DISCUSSION

Strategy of Genetic Mapping by Restriction Enzyme Site Polymorphisms. This approach has three basic steps. First, single- or low-copy probes are isolated from each of the 13 class I gene clusters. Second, each of these probes is used to analyze the DNAs of various mouse inbred and recombinant strains for restriction enzyme site polymorphisms. Third, these restriction enzyme site polymorphisms are correlated with the serologic class I polymorphisms that define the four regions (*K*, *D*, *Qa*, and *Tla*) in the MHC by recombinational analysis. Thus, the cosmid cluster positions are identified by the correlation of the polymorphisms in single- or low-copy probes and class I serological markers.

Single- and low-copy probes isolated from the 13 cosmid clusters are listed in Table 1. These probes fell into two categories. Three probes were single copy—those from clusters 4, 7, and 10—and the 10 low-copy probes from the remaining clusters each hybridized with two to eight restriction fragments under stringent conditions. In those cases in which the low-copy probes hybridized to multiple bands from BALB/c DNA, the parental band corresponding to the probe was identified by virtue of its comigration with cloned BALB/cDNA, because it was the most intense band in the autoradiogram, or both.

In theory, an analysis of the single-copy probes against BALB/c DNA and the genomic DNAs from various mouse inbred strains might show one of four types of hybridization patterns. These patterns are (*i*) no change in band size (no restriction enzyme site polymorphism), (*ii*) a change in band size (mutation of a restriction enzyme site), (*iii*) loss of the band (deletion of the probe sequence), and (*iv*) multiple bands (duplication of the probe sequence). The last three patterns can be used as genetic markers to map the corresponding sequences to positions in the MHC. Each of these patterns was observed in analyzing various mouse DNAs with our single- and low-copy probes.

The final step in mapping by this approach is to correlate the restriction enzyme site polymorphisms of various DNAs with the serologically defined recombinational maps of the corresponding congeneic and recombinant congeneic mouse strains.

Five Class I Genes from Three Clusters Map to the *K* and *D* Regions of the *H-2* Complex. Because a useful combination of restriction enzyme site polymorphisms and recombinant congeneic strains of mice is not always available, it often is necessary to go through a multistep analysis to map the cosmid clusters. The mapping of cluster 13 to the *D* region required three independent analyses. First, the 3.4-kilobase (kb) fragment used as a low-copy probe detects five restriction enzyme fragments in *Eco*RI-digested BALB/c DNA. The 25-kb fragment is the most intense and therefore presumably contains the probe sequence (Fig. 1). This band is missing in AKR DNA and in the DNA of the congeneic mouse strain B6-H-2^k. These initial observations map cluster 13 to the MHC of the mouse, because B6-H-2^k mice contain the MHC from AKR mice on the B6 background and the DNA from B6 mice contains the 25-kb fragment. Second, an analysis of *Hin*dIII-digested genomic DNA using the same probe shows that the 13.5-kb fragment detected in BALB/c, A/WySn, and B10.A(3R) DNAs is replaced by a 10.6-kb fragment in C57BL/6 DNA. These observations map the probe for cluster 13 distal in the recombination point in the B10.A(3R), which is located in the class II region between the *K* and *D* loci (Table 1). Finally, because the 25-kb *Eco*RI fragment that is absent from the DNA from B6-H-2^k is present in

recombinant congeneic mouse strains

Clusters, kb							
						13	
7 (*Pvu* II)	8 (*Hin*dIII)	9 (*Hin*dIII)	10 (*Bam*HI)	11 (*Hin*dIII)	12 (*Hin*dIII)	*Eco*RI	*Hin*dIII
2.8	3.2	12.3	9	10.6	3.4	25	13.5
1.6	7.7	8.8	9	10.6	—		10.6
				10.6		12.3	
2.8	3.2	8.4	9	5.6	—	25	10.6
2.8			9				
1.6	—		11		—		13.5
			9				
			9		3.4		13.5
			11				
1.6	—		9				
2.8	—				—		
	3.2		9	10.6	—		
	—	12.3	11	5.6	—		13.5
				10.6			10.6
	3.2				3.4		
			9		—		
2.8	—	8.4	11		—	25	
2.8	7.7	8.8			—	12.3	
2.8	7.7	8.8			—	25	
2.8	7.7	8.4			—	25	
Tla	*Tla*	*Qa*	*Tla*	*K*	*Qa/Tla*	*D*	

site of M13 mp8; cluster 10, 3-kb *Xho* I fragment of cosmid 15.3; cluster 11, 2.6-kb *Xho* I/*Nru* I fragment of cosmid 17.1 subcloned into the *Hin*cII site of M13 mp8; cluster 12, 2-kb *Bam*HI fragment of cosmid 22.1 subcloned into the *Bam*HI site of M13 mp8; cluster 13, 3.4-kb *Sma* I/*Nru* I fragment of cosmid 18.1. MB, Because of multiple bands, the corresponding fragment could not be identified.

*Taken from refs. 15 and 16.

†The corresponding fragment could not be identified because of deletion or comigration with other bands.

the recombinant congeneic strain B6.K1, which is identical with the B6-H-2^k strain at both the *Tla* and *Qa-2,3* loci but differs at the *K* and *D* loci, we can conclude cluster 13 must map proximal to the *Tla* complex. Together, these restriction enzyme site polymorphisms suggest that cluster 13 maps to the *D* region. Similar approaches have mapped cosmid clusters 11 and 2, respectively, to the *K* and *D* regions (Table 1).

Ten Class I Genes from Three Clusters Map to the *Qa* Region. Clusters 1, 6, and 9, which contain seven, two, and one class I genes, respectively (9), map to the *Qa* region (Table 1). We will discuss the data mapping cluster 6 as an example. The low-copy probe isolated from cluster 6 detected a 10.5-kb *Kpn* I fragment in BALB/c (*d* haplotype) DNA whereas a 16-kb and a 13-kb *Kpn* I fragment were detected in C3H (*k* haplotype) and in C57BL/6 (*b* haplotype) DNAs, respectively. Since the recombinant congeneic strain B10.A(1R) has the *b* haplotype-specific fragment, we can localize this cluster distal to the *S* region. Analyses of the strains B6-Tla^a (13-kb fragment), B6-H-2^k (16-kb fragment), and their recombinant B6.K2 (13-kb fragment) allow us to place cluster 6 proximal to the *Tla* region because B6.K2 DNA contains the *Tla* region genes from the B6-H-2^k mouse. The recombinant congeneic strain B6.K1, also derived by recombination from B6-Tla^a and B6-H-2^k, however, showed an interesting result in that both the 16- and the 13-kb *Kpn* I fragments were found. The simplest explanation of this finding is that the B6.K1 haplotype was generated by an unequal crossing-over event between B6-Tla^a and B6-H-2^k chromosomes, leading to a duplication of DNA at the site of recombination. Cluster 6 would map close to this recombination point and has therefore been duplicated in strain B6.K1. Cluster 9 can be mapped only tentatively to the *Qa* region because multiple restriction fragments hybridized with the probe from this cluster.

Fourteen Class I Genes from Five Clusters Map to the *Tla* Region. Clusters 3, 4, 7, 8, and 10, which have six, three, two, two, and one class I genes, respectively (9), were mapped to the *Tla* region (Table 1). Fig. 1 shows the results obtained with a single-copy probe isolated from cluster 4. This probe identifies a 14-kb *Kpn* I fragment in BALB/c and C57BL/6 DNA whereas this sequence is absent from strain A/WySn DNA. Analysis of the congeneic strain B6-Tla^a (*Tla* region from A on a C57BL/6 background) shows that this sequence is absent from this strain also and therefore maps cluster 4 to the *Tla* region. As a control, Fig. 1 shows that the 14-kb fragment is present in strains B6.K1, B6.K2, and B10.A(1R) (*Tla* regions from AKR and C57BL/10) but absent from strains B10.A(3R) and B10.A (Tla region from A).

Seven Class I Genes from Two Clusters Map to Either the *Qa* or the *Tla* Region. Clusters 5 and 12, which have four and three class I genes, respectively (9), map in the MHC distal to the *D* region (Table 1). Due to the lack of appropriate strains and restriction enzyme site polymorphisms, it is not possible to map these clusters more precisely with the probes used. From cluster 5, a 2.6-kb fragment was isolated and used to identify a 13-kb *Kpn* I fragment in BALB/c DNA from which this low-copy probe was derived. As shown in Fig. 1, this 13-kb *Kpn* I fragment is replaced by a 10-kb *Kpn* I fragment in DNA from C57BL/6 and B10.BDR-1 mice. All other strains analyzed show the 13-kb *Kpn* I fragment. Because the congeneic and recombinant congeneic strains B6-Tla^a, B6.K1, and B6.K2 do not show the C57BL/6-specific 10-kb band, we can conclude that cluster 5 maps to the right of the *D* locus on chromosome 17. A correlation with the *Qa-2,3* and *Tla* alleles present in these strains (Table 1) is not possible, indicating that cluster 5 maps to a locus in the *Tla* complex not yet defined by serological reagents.

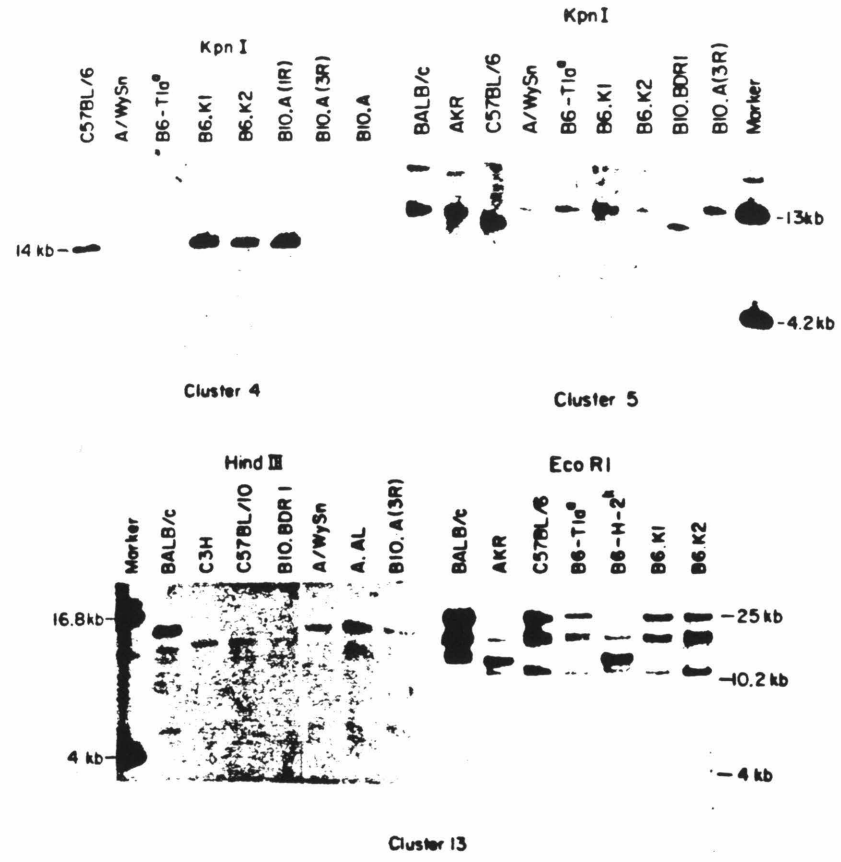


FIG. 1. Southern blot analyses of DNAs from inbred, congeneic, and recombinant congeneic mice with the single- or low-copy probes isolated from clusters 4, 5, and 13.

A Class I Gene Cluster Possibly Mapping to the Distal End of the *Tla* Region. Comparison of the restriction maps for clusters 7 and 8 shows that position 2 (27 kb) of cluster 7 overlaps position 12 (37 kb) of cluster 8 (9). In fact, these two clusters can be merged into one cluster of 70 kb containing three class I genes if it is assumed that part of cosmid 8.3 in cluster 8 [position 37 (47 kb) (9)] is derived from noncontiguous DNA ligated to the class I gene sequence during construction of the library. This possible interpretation escaped our notice earlier (9). The low-copy probe isolated from the overlapping portion between clusters 7 and 8 (3-kb *Hpa* I fragment) map these class I genes to the *Qa* and *Tla* regions (Table 1). A single-copy probe isolated further downstream from the nonoverlapping portion of cluster 7 (a 3-kb *Xho* I fragment), however, maps outside the MHC (Table 1). Assuming that this *Xho* I end of cluster 7 (which is defined only by a single cosmid clone) does not represent a cloning artifact, this would map cluster 7 to the distal end of the *Tla* region encompassing the sequence containing the recombination point in a number of congeneic mouse strains. Indeed, a second low-copy probe from cluster 7 (a 1.5-kb *Bst*EII/*Nru* I fragment) together with the 3-kb *Xho* I fragment as probe allowed us to establish, by restriction enzyme analysis of the genomic DNA, that the organization of the cloned DNA in cluster 7 is colinear with genomic DNA (data not shown).

Duplication and Deletion of Class I Gene Clusters in Inbred Mice. Southern blot analyses with class I cDNA probes have shown that different inbred strains of mice contain similar numbers of class I genes (12, 17–20). With the probes that we have isolated from the 13 gene clusters, we are now in a position to analyze the duplication and deletion of class I gene clusters in a more precise way by counting the total numbers of bands that are generated by Southern blot analyses of DNAs from various inbred strains. The assumption made here is that the duplication or deletion of the probe sequences is roughly proportional to the duplication and deletion of adjacent DNA sequences such as the class I genes. We have previously reported an example of DNA deletion in that a sequence for the 5′ flanking part of the L^d gene was deleted in strains of the *b* and *k* haplotypes, indicating a possible deletion of the L^d coding sequence, which is not expressed in these two strains (9). Furthermore, as discussed above, a duplication of the sequence used as a probe for cluster 6 occurs in recombinant strain B6.K1. Using all 13 low- or single-copy probes, we have detected 124 bands in BALB/c DNA, 102 bands in C3H DNA, and 107 bands in C57BL/6 DNA. Thus, expansion and contraction of the probe sequences and presumably the class I gene does occur in these different inbred strains.

All 36 Class I Genes Map to the MHC. A striking finding is that all 36 of the isolated class I genes map to the MHC. In several other multigene families, including globin (21) and the antibody genes (22, 23), it appears that pseudogenes can map to other chromosomal locations than the functional genes. We cannot formally exclude the possibility that there are additional class I genes or distantly related pseudogenes not present among the cloned cosmid genes that lie outside the MHC.

Five of the 36 genes map to the classical *H-2* complex (Fig. 2). Three of these five genes have been shown by DNA-mediated gene transfer to express the *K*, *D*, and *L* gene products.

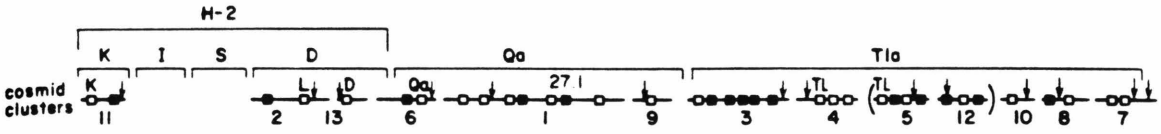


FIG. 2. Locations of the 13 class I gene clusters with respect to the genetic map of the major histocompatibility complex. Cluster 11 maps to the *K* region; clusters 2 and 13 map to the *D* region; clusters 1, 6, and 9 map to the *Qa* region; clusters 3, 4, 7, 8, and 10 map to the *Tla* region; and clusters 5 and 12 map to either the *Qa* or the *Tla* region. The order of clusters in separate regions is not known with respect to each other. The locations of the probes used for mapping are indicated by arrows. The genes in each cluster are denoted by boxes. Open boxes signify elevated β_2-microglobulin levels whereas genes given by closed boxes did not lead to elevated levels of surface β_2-microglobulin in gene-transfer experiments. The implication is that the former class I genes are expressed and the latter are not by this assay (10).

The other two genes fail to express gene products associated with β_2-microglobulin on the cell surface (10) and are provisionally assumed to be pseudogenes. Thus, the *H-2* complex encoding the classical transplantation antigens appears to contain a very modest number of class I genes as compared with the *Tla* complex encoding the structurally related hematopoietic differentiation antigens. Thirty-one of the 36 class I genes map to the *Qa* and *Tla* regions (Fig. 2). At least three of these class I genes express the serologically defined *Qa-2,3* and two *TL* gene products (10). In addition, at least 10 additional class I genes in the *Tla* complex express what we have termed as novel class I gene products (10). The novel gene products were detected by increased levels of β_2-microglobulin on the surface of mouse L cells that had been transformed with these class I genes.

Mapping the class I genes by restriction enzyme site polymorphisms assumes that the flanking sequences from which we isolate the probes are not rearranged with respect to the class I genes in the other mouse strains analyzed. Three observations support this assumption. First, cluster 1 has been mapped independently to the *Qa* region with two probes, gene *27.1* (9), and the 2.1-kb restriction fragment from the 3′ flanking sequence of gene *2* (this paper). Thus, two sequences linked by molecular cloning are also linked by genetic mapping. Second, a precise correlation exists between the results obtained by serological identification of class I genes and the mapping of the corresponding cluster by restriction enzyme polymorphism with flanking sequence probes. Third, probes isolated from the clusters mapping to the *Qa* and *Tla* regions detected the same sized restriction fragments in multiple mouse strains, indicating a general conservation of sequence organization in these regions.

Restriction Enzyme Site Polymorphisms Correlate with Serological Polymorphisms. In agreement with the extreme polymorphism of the transplantation antigens is our finding that restriction enzyme site polymorphisms were readily detected with probes from the cosmid cluster containing the *K*, *D*, and *L* genes. This indicates that the allelic variability of the coding regions of the *K*, *D*, and *L* molecules as defined serologically does extend into the flanking DNA sequences. For the class I gene clusters mapping to the *Tla* complex, it was much more difficult to detect restriction enzyme site polymorphisms. For example, 10 restriction enzymes were analyzed for cluster 10 before a polymorphism was found. Once again, the limited restriction enzyme site polymorphisms correlate with the limited serological polymorphism of the Qa-1, Qa-2,3, and TL antigens. Thus, certain regions of the major histocompatibility complex appear to exhibit extensive polymorphism (*K* and *D*) whereas others do not (*Qa* and *Tla*). The explanation for these striking differences is unknown. This difference in polymorphisms between the numerous class I genes of the *Tla* complex and the few class I genes of the classical *H-2* complex also explains the limited restriction enzyme polymorphism that is detected in Southern blot analyses of different mouse DNAs with class I gene sequences (12, 17–20).

We thank Karyl Minard for excellent technical help, Mary J. Macchi and Dr. Jeff Frelinger for some of the recombinant congeneic mouse DNAs used in the Southern blot analyses, and Bernita Larsh for the preparation of the manuscript. We also thank Henry Sun, Drs. Steve Crews, Joan Kobori, Marie Malissen, Bernard Malissen, Roger Perlmutter, and Martha Zuniga for comments on the manuscript. A.W. was supported by a fellowship from the California Foundation for Biochemical Research. M.S. was supported by a Senior Leivre Fellowship from the California Division of the American Cancer Society. The work was supported by grants from the National Institutes of Health.

1. Klein, J. (1975) *Biology of the Mouse Histocompatibility Complex* (Springer, Berlin).
2. Snell, G. D., Dausset, J. & Nathenson, S. (1976) *Histocompatibility* (Academic, New York).
3. Nathenson, S. G., Uehara, H., Ewenstein, B. M., Kindt, T. J. & Coligan, J. E. (1981) *Annu. Rev. Biochem.* **50**, 1025–1052.
4. Ploegh, H. L., Orr, H. T. & Strominger, J. L. (1981) *Cell* **24**, 287–299.
5. Klein, J. (1979) *Science* **203**, 516–521.
6. Zinkernagel, R. M. & Doherty, P. C. (1980) *Adv. Immunol.* **27**, 51–177.
7. Flaherty, L. (1980) in *The Role of the Major Histocompatibility Complex in Immunology*, ed. Dorf, M. E. (Garland, New York), pp. 33–57.
8. Hood, L., Steinmetz, M. & Goodenow, R. (1982) *Cell* **28**, 685–687.
9. Steinmetz, M., Winoto, A., Minard, K. & Hood, L. (1982) *Cell* **28**, 489–498.
10. Goodenow, R. S., McMillan, M., Nicolson, M., Sher, B. T., Eakle, K., Davidson, N. & Hood, L. (1982) *Nature (London)* **300**, 231–237.
11. Steinmetz, M., Zachau, H. G. & Mach, B. (1979) *Nucleic Acids Res.* **6**, 3213–3229.
12. Steinmetz, M., Moore, K. W., Frelinger, J. G., Sher, B. T., Shen, F. W., Boyse, E. A. & Hood, L. (1981) *Cell* **25**, 683–692.
13. Steinmetz, M., Minard, K., Horvath, S., McNicholas, J., Frelinger, J., Wake, C., Long, E., Mach, B. & Hood, L. (1982) *Nature (London)* **300**, 35–42.
14. Steinmetz, M., Höchtl, J., Schnell, H., Gebhard, W. & Zachau, H. G. (1980) *Nucleic Acids Res.* **8**, 1721–1729.
15. Klein, J., Flaherty, L., Van de Berg, J. L. & Shreffler, D. C. (1978) *Immunogenetics* **6**, 489–512.
16. Klein, J., Figueroa, F. & Klein, D. (1982) *Immunogenetics* **16**, 285–317.
17. Steinmetz, M., Frelinger, J. G., Fisher, D., Hunkapiller, T., Pereira, D., Weissman, S. M., Uehara, H., Nathenson, S. & Hood, L. (1981) *Cell* **24**, 125–134.
18. Cami, B., Bregegere, F., Abastado, J. P. & Kourilsky, P. (1981) *Nature (London)* **291**, 673–675.
19. Margulies, D. H., Evans, G. A., Flaherty, L. & Seidman, J. G. (1982) *Nature (London)* **295**, 168–170.
20. Pease, L. R., Nathenson, S. G. & Leinwand, L. A. (1982) *Nature (London)* **298**, 382–385.
21. Leder, A., Swan, D., Ruddle, F., D'Eustachio, P. & Leder, P. (1981) *Nature (London)* **293**, 196–200.
22. Hollis, G. F., Hieter, P. A., McBride, O. W., Swan, D. & Leder, P. (1982) *Nature (London)* **297**, 83–84.
23. Battey, J., Max, E. E., McBride, O. W., Swan, D. & Leder, P. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 5956–5960.

Chapter Four

GENOMIC ORGANIZATION OF THE GENES ENCODING MOUSE T-CELL RECEPTOR α-CHAIN

Published in

Nature **316**, 832-836, 1985

Reprinted from Nature, Vol. 316, No. 6031, pp. 832-836, 29 August 1985

Genomic organization of the genes encoding mouse T-cell receptor α-chain

Astar Winoto, Shelley Mjolsness & Leroy Hood

Division of Biology, California Institute of Technology, Pasadena, California 91125, USA

The vertebrate immune system uses two kinds of antigen-specific receptors, the immunoglobulin molecules of B cells and the antigen receptors of T cells. T-cell receptors are formed by a combination of two different polypeptide chains, α and β (refs 1–3). Three related gene families are expressed in T cells, those encoding the T-cell receptor, α and β, and a third, γ (refs 4–6), whose function is unknown. Each of these polypeptide chains can be divided into variable (V) and constant (C) regions. The V_β regions are encoded by V_β, diversity (D_β) and joining (J_β) gene segments that rearrange in the differentiating T cell to generate V_β genes[7–14]. The V_γ regions are encoded by V_γ, J_γ and, possibly, D_γ gene segments[4,5]. Studies of α complementary DNA clones suggest that α-polypeptides have V_α and C_α regions and are encoded by V_α and J_α gene segments and a C_α gene[15–17]. Elsewhere in this issue we demonstrate that 18 of 19 J_α sequences examined are distinct[18], indicating that the J_α gene segment repertoire is much larger than those of the immunoglobulin (4–5) or β (14) gene families. Here we report the germline structures of one V_α and six J_α mouse gene segments and demonstrate that the structures of the V_α and J_α gene segments and the α-recognition sequences for DNA rearrangement are similar to those of their immunoglobulin and β-chain counterparts. We also show that the J_α gene-segment organization is strikingly different from that of the other immunoglobulin and rearranging T-cell gene families. Eighteen J_α gene segments map over 60 kilobases (kb) of DNA 5′ to the C_α gene.

A full-length cDNA probe, A10, was obtained for the isolation of germline V_α, J_α and C_α clones from a λgt10 cDNA library of the T-cell helper hybridoma 1.9.2 using an oligonucleotide probe specific for the C_α gene[15,16]. DNA sequence analysis showed that the A10 clone was identical in sequence to the published full-length cDNA clone TT11 derived from a T-cell hybridoma[15]. A V_α-specific probe was derived from the A10 cDNA and used in Southern blot analysis to demonstrate a V_α rearrangement in the BW5147 tumour DNA (data not shown). As both α-sequences come from different T-cell hybridomas which have the BW5147 cell line as a common parent, these observations suggest that clone A10 is derived from the T-cell tumour BW5147 α-gene. This V_α-specific probe was then used to isolate both germline and rearranged clones of the BW5147 V_α subfamily from a λ genomic library of the 1.9.2 hybridoma DNA.

A C_α-specific probe was obtained from the A10 cDNA clone and used in Southern blot analysis of several mouse DNA clones digested with various restriction enzymes. These data indicate that there is only one cross-hybridizing C_α gene (data not shown). The C_α-specific probe was then used to screen a cosmid library constructed from liver DNA of the inbred B10.D2-H-2^{dm1} mouse to obtain three overlapping cosmid clones (Fig. 1). Two successive chromosomal walks were then performed with single-copy probes isolated from the 5′ ends of cosmid clones TA4.1 and TA52.1 to yield five additional clones (Fig. 1). These eight cosmid clones encompassed 119 kb of DNA and contained all 18 J_α gene segments described elsewhere in this issue[18].

To study the structure of the V_α gene segment, we isolated a germline V_α clone, V_α5H, using the V_α-specific probe from the A10 cDNA. The V_α5H has a leader exon of 49 nucleotides separated by an intron of 188 nucleotides from a V exon of 287 nucleotides (Fig. 2*a*). The leader coding segment is split by an intron probably four codons from its 3′ end. The V_α5H sequence seems to represent a functional V_α gene segment and belongs to the V_α1 subfamily which contains at least 10 different members (ref. 18 and A.W., unpublished data). The V_α5H coding region is 96% homologous at the DNA level and 98% homologous at the protein level to the coding region of the A10 cDNA clone, hence these sequences represent closely related members of the V_α1 subfamily (see ref. 18).

The 3′ end of the V_α gene segment is marked by a recognition sequence for DNA rearrangement[19]. It consists of a conserved heptamer sequence CACAGTG, a 21-base pair (bp) spacer sequence and an A+T-rich nonamer sequence (Fig. 2). The recognition signal is very similar to those of the κ and heavy (H) immunoglobulin[19] and β and γ T-cell gene families[5,7,8,10,11].

The C_α gene is 79 kb from the 5′ end of the cosmid cluster and 37 kb from the 3′ end (Fig. 1 and A.W., unpublished data).

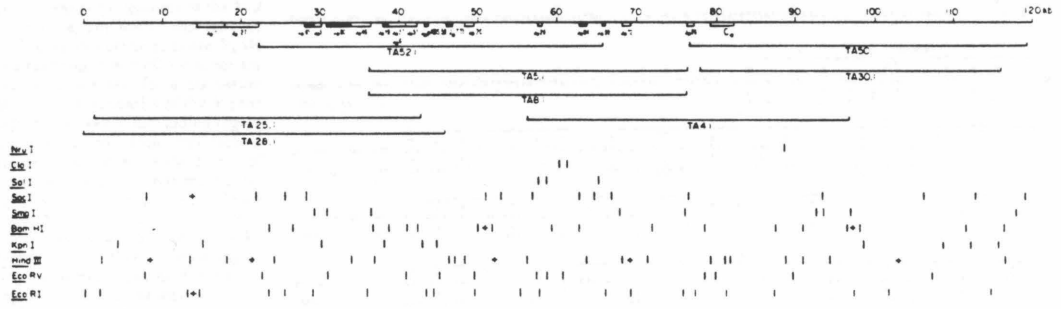


Fig. 1 Restriction map of the mouse cosmid cluster containing the J_α gene segments and the C_α gene. Open boxes indicate restriction fragments that hybridize to the 500-bp *Sau*3A fragment C_α probe from A10 cDNA clone or to the synthetic oligonucleotide J_α probes. Solid boxes indicate the J_α gene segments that have been sequenced in this paper. Unassigned sites between two restriction sites are indicated by +. The J_α gene segments lie to the 5′ side of the C_α gene whose direction of transcription is to the right. The J_α gene segments are named according to the cDNA clone from which they originated[15,16,18]. The cDNA library was constructed in λgt10 vector as described[38], with minor modification. The second-strand synthesis was done using RNase H, DNA polymerase I and T4 ligase[39] and the unincorporated nucleotides were removed at every step by NH_4-acetate/ethanol precipitation. The cosmid library was constructed and screened as described previously[40]. The oligonucleotides for J_α gene segments were purified in a 20% preparative polyacrylamide gel. The DNA band was visualized by placing a silica gel 60F-254 precoated TLC plate (EM reagents) under the gel and shining short ultraviolet light from above. The DNA band was then cut out, crushed and mixed with 3 ml 0.5 M NH_4-acetate, 10 mM Mg-acetate, 1 mM EDTA, 0.1% SDS and elution was at 37 °C for 3 h with shaking. The acrylamide was then separated from the DNA by spinning in a Quick-sep Isolab column and the aqueous solution containing oligonucleotide was concentrated to <0.2 ml by several *sec*-butanol extractions. Desalting of DNA was done in a 5 ml Sephadex G25 column. Fractions were collected and monitored by a spectrophotometer. The pure oligonucleotides were kinased as described previously[41]. Hybridization of the J_α oligonucleotides was done in 6×SSC, 5×Denhardt's, 0.1% SDS and 100 μg ml^{-1} denatured salmon sperm DNA at 37 °C with 5×10^5 c.p.m. ml^{-1}. The blots for oligonucleotides were washed at 25 °C in 6×SSC, 0.1% SDS.

The locations of J_α gene segments were determined using 18 oligonucleotide probes that were synthesized corresponding to the 5′ or 3′ ends of the J_α cDNA sequences[15,16,18]. As the 5′ and 3′ ends of the J_α coding regions are different from one another[18], the oligonucleotide probes do not cross-hybridize in the conditions used. In each case only one restriction enzyme fragment from the cosmid clones hybridized to each of the individual oligonucleotide probes. Eighteen J_α gene segments map over a region which extends from 3 to 63 kb 5′ to the C_α gene (Fig. 1). The precision in the assignment of the J_α gene segment locations varies within 0.5 to 4 kb of DNA because of the limited number of restriction enzymes that we have used to map the cosmid clones. Hence, the distribution of J_α gene segments is strikingly different from the distribution of the J gene segments of immunoglobulin and the T-cell β- and γ-genes. The J_α gene segments are spread over at least 60 kb of DNA, whereas the J_β and J_H gene clusters extend over only 2 kb of DNA. Assuming that the 18 J_α gene segments analysed are a random representation of the J_α gene segment repertoire, the fact that all of them map to the available cosmid clones is consistent with the hypothesis that most of the J_α gene segments will map to the 63-kb region of DNA 5′ of the C_α gene. As only one of 19 J_α sequences analysed in the accompanying paper[18] was a repeat, the repertoire of J_α sequences is probably larger than 18.

Six germline J_α gene segments were subcloned into M13 bacteriophage and sequenced using specific oligonucleotide primers[20] (Fig. 3). The coding regions of the J_α sequences were identified by a sequence encoding Phe-Gly-X-Gly that is conserved in virtually all T-cell J gene segments. The 5′ boundary of the J_α gene segments was determined by the presence of a one-turn recognition sequence for DNA rearrangement. The 3′ boundary was determined by analyses of the germline and cDNA sequences[18] in this region which permitted identification of the site for RNA splicing between the J_α and C_α sequences.

These are several unique features of the germline J_α segments. First, they are longer than the J gene segments of immunoglobulin and β-genes by 3–9 codons. Second, the J_α gene segments exhibit length heterogeneity at their 3′ ends; this heterogeneity is in contrast to the other J gene segments within a particular family (λ, κ, H, β), which all have the same lengths at their 3′ ends. All the immunoglobulin and T-cell J gene segments exhibit length heterogeneity at their 5′ ends. Thus, the J_α sequences are more heterogeneous than their immunoglobulin and β counterparts[18].

The recognition sequences for DNA rearrangement that lie 5′ of each of the J_α gene segments are very similar to those found in the immunoglobulin and β-gene families (Fig. 3). They each have a highly conserved heptamer separated from an A+T-rich nonamer by a 12-bp spacer sequence. In addition, the 3′ splicing site for joining the J_α to the C_α RNA sequences contains the canonical RNA splicing donor sequence GTAAG[21].

The fact that the V_α gene segments have recognition sequences with 23-bp spacers (two DNA helix turns) and the J_α gene segments have recognition sequences with 12-bp spacers (one turn) establishes that all three rearranging gene families in T cells (α, β and γ) use similar kinds of recongition sequences for their V and J gene segments. In the case of the β-genes, the D_β gene segments have a one-turn recognition sequence to the 5′ end and a two-turn recognition sequence to the 3′ end and accordingly join one- to two-turn recognition sequences in the process of rearrangement and juxtaposition of the V_β, D_β and J_β gene segments[9]. The question of whether or not D_α gene segments exist remains. However, even without D_α gene segments, a V_α may join to the J_α gene segment because of the complementary one- and two-turn recognition sequences.

The J_α gene segments are more widely spaced from one another than are their immunoglobulin and β-gene counterparts. The immunoglobulin and J_β gene segments within a cluster are separated by 36–519 nucleotides, with most of the J gene segments separated by <350 bp. We have sequenced 500 and 400 bp to the 5′ and 3′ side of the J_α19 gene segment, 400 and 500 bp to the 5′ and 3′ sides of J_α80, 400 bp to the 5′ side of J_α65 and 200 bp on both sides of the J_αTT11 gene segment. In addition, we have sequenced more than 900 bp 5′ to the J_α pHDS 58 gene

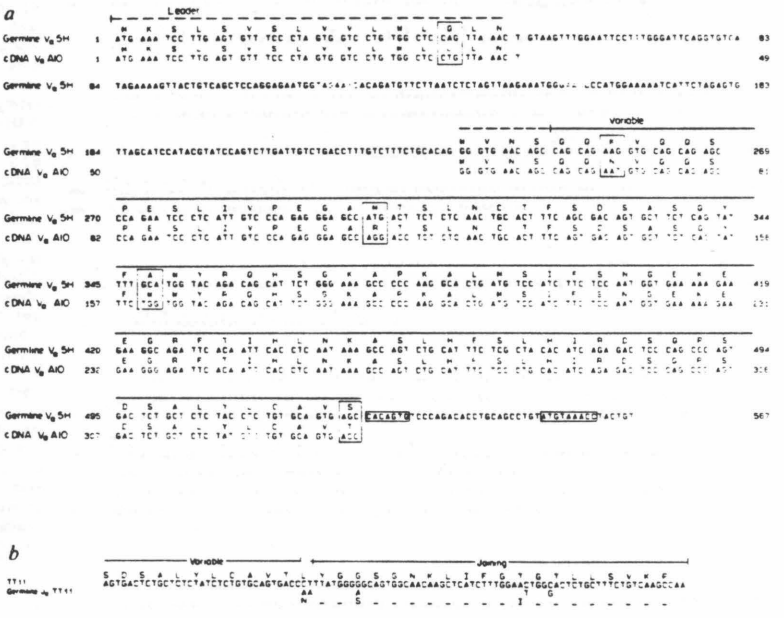


Fig. 2 *a*, DNA sequence of the germline and the rearranged V_α gene segments; *b*, alignment of germline and rearranged sequences of the *V–J* junction. *a*, The leader and variable coding regions of the germline V_α5H and rearranged A10 cDNA clone are indicated. See text for a discussion of 5′ and 3′ boundaries of the *V* gene segment. Boxes in the coding region indicate codons with replacement changes and boxes in the 3′ end of the V_α gene segment indicate the recognition sequence for DNA recombination. The A10 cDNA clone was obtained from the 1.9.2 helper T-cell cDNA library using an oligonucleotide specific for the C_α region[15,16]. The clone was subcloned in M13mp18 and sequenced by the dideoxy chain termination method as described elsewhere[20,42]. The λ genomic library from the 1.9.2 helper T-cell hybridoma was constructed in EMBL 3 vector as described[41,43]. The library was screened using a 300-bp *Rsa*1/*Eco*RI V_α-specific probe from A10 cDNA. The V_α-containing fragment from the V_α5H λ clone was subcloned into M13mp18 or mp19 and sequenced as described[20,42]. *b*, Upper line represents the TT11 cDNA sequence and the lower the germline J_α sequence. Above and below the nucleotide sequences are the translated protein sequences. *V* and *J* regions are indicated. Dots between the *V* and *J* region indicate extra nucleotides between them; dots below the nucleotide sequences indicate identity to the DNA sequence given above; a dash indicates identity to the protein sequence shown above. Only the 3′ part of the V_α region of the cDNA is shown.

segment (Fig. 3 and A.W., unpublished data); in each case, we failed to observe an adjacent J_α coding region as defined by the invariant Phe-Gly-X-Gly sequence.

A comparison of homologous sequences from the J_α gene segments isolated from DNA of the B10.D2-H-2^{dm1} mouse and the cDNA clones derived from AKR and BALB/c mice[15,16] shows that these J_α sequences have several nucleotide differences as well as a one-codon sequence gap. Because each J_α oligonucleotide probe hybridizes to just a single germline sequence, these observations suggest that J_α gene segments are polymorphic. Analysis of V_α gene segments also reveals that there are many polymorphisms in the V_α gene segments[18]. The polymorphism that is present in the α-gene family is significantly greater than that seen in the β-gene family[22].

The presence of a D_α gene segment in a rearranged V_α gene should lead to additional nucleotide sequence between the germline V_α and J_α gene-segment sequences. The alignment of the A10 cDNA clones and the available germline V_α and J_α gene segments demonstrates that there are two extra nucleotides between the V_α and J_α gene segments in this clone (Fig. 2*b*). The J_α gene segment is the germline equivalent to that expressed in the A10 cDNA, whereas the germline V_α gene segment belongs to the same V_α subfamily but is not identical. In the V_κ, V_H and V_β gene families, members of the same V subfamily always end at the same codon position relative to the conserved second cysteine codon. The additional nucleotides between the rearranged V_α and J_α sequences may be explained in three ways. First, the extra nucleotides can arise from D_α gene segments that are joined to the V_α and J_α gene segments. Second, these extra sequences could arise as a result of N-region diversification. This mechanism results in the addition of random nucleotides to the 5′ and/or 3′ sides of the D gene segments during the joining process[23]. In the immunoglobulin light-chain gene families, which do not have D gene segments, N-region diversity does not occur. Third, it is possible that the extra nucleotides are actually part of the V_α gene segment. These three possibilities cannot yet be distinguished. However, either D_α gene segments exist, N-region diversification occurs in a rearranging gene family lacking D gene segments, or the V_α gene segments within a subfamily differ in the lengths of their 3′ ends.

The immunoglobulin and β-gene families use various mechanisms for somatic diversification. Junction variability occurs as a consequence of the flexibility in sites at which the gene segments are joined; and somatic hypermutation, the apparently random mutation of up to 3% of the bases in and around a rearranged V gene, occurs in immunoglobulin genes, but rarely in β-genes[24,25].

The α-gene family also uses various mechanisms for somatic diversification. An alignment of the germline V_α and J_α gene

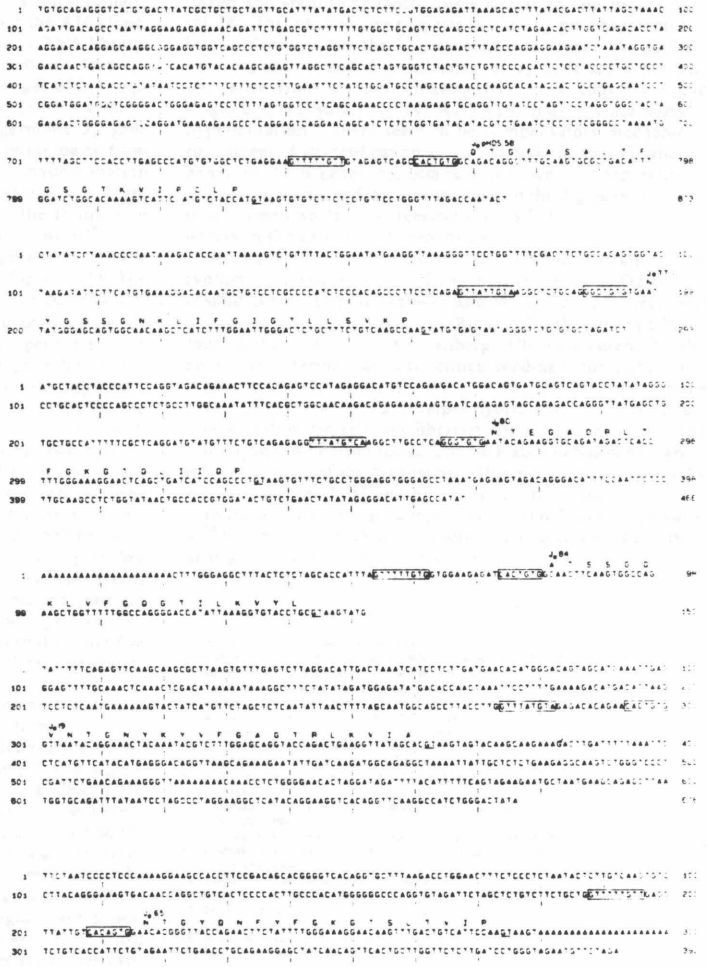


Fig. 3 DNA sequence of six germline J_α gene segments. The J gene segments are translated into protein sequence with the one-letter code. The J_α designations are given above the protein sequence. Recognition sequences for recombination are boxed and the RNA donor splicing sites are underlined. Cosmid restriction fragments containing the J_α gene segments were purified using low-melting point agarose and then subcloned into M13 bacteriphage mp18 or mp19 or both[44]. Plaques of M13 containing the J_α sequences in both orientations were identified by hybridization to the J_α oligonucleotides made from DNA sequences of both strands. The following cosmid fragments were used for subcloning: 0.5-kb *Hin*dIII fragment for J_αTT11, 1.8-kb *Kpn*I fragment for J_αpHDS 58, 2-kb *Xba*I fragment for J_α80, 1-kb *Bam*HI/*Hin*dIII fragment for J_α84, 2-kb *Bam*HI fragment for J_α19 and 0.5-kb *Xba*I fragment for J_α65. DNA sequencing by the dideoxy chain termination method was done essentially as described previously[42] and modified[20]. Some of the DNA sequencing was done by the chemical degradation method[45]. DNA sequencing was done for both strands. For every J_α-containing fragment >0.5 kb, two complementary 17-mer J_α oligonucleotides were used for the first strand sequence analysis. The opposite strands of DNA were sequenced by two additional oligonucleotides (17-mer) made according to the information obtained in the first strand DNA sequencing. Hence, at least four different oligonucleotides were made and used for sequencing one J_α gene segment to obtain information on both strands. In one case in which we could not obtain clones containing opposite orientations of the J_α gene segment (J_αTT11), the sequence was confirmed by the Maxam–Gilbert sequencing method[45].

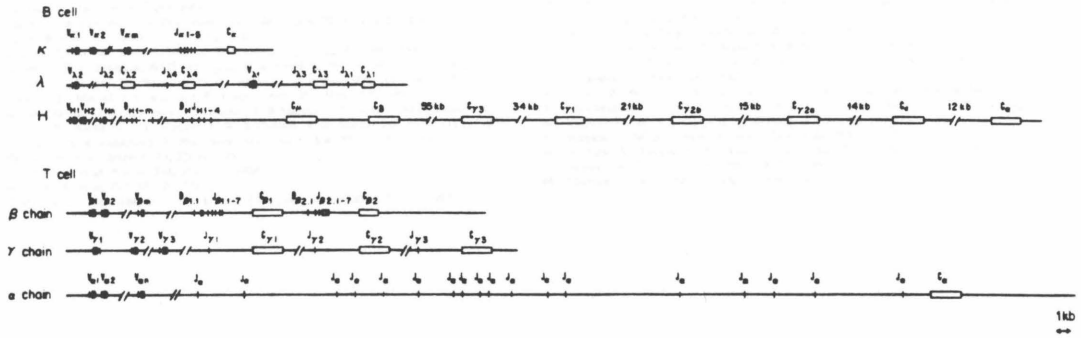


Fig. 4 Schematic diagram of the gene organization for the immunoglobulin and T-cell receptor gene families. Distances between J and C gene segments for immunoglobulin[27,28,31,46] and T-cell rearranging[5,7,10,11] families are indicated. Boxes indicate the V gene segments or the C genes, vertical lines the D or the J gene segments.

segments with their cDNA counterparts revealed that there is flexibility in the sites at which the gene segments are joined together. For example, in the TA19 cDNA[18], the J_α gene segment is joined five bases into the coding sequence. Thus, diversity can be introduced in the junctional region by a mechanism similar to those seen in the immunoglobulin and β-gene families. The data are inadequate to draw firm conclusions about somatic hypermutation in V_α genes because, apart from the A10 clone, these cDNA clones were derived from a thymic cDNA library. In B cells somatic hypermutation occurs late in B-cell development[26] and T cells in the thymus are at early stages of development. Two sets of α-gene segments do not exhibit somatic mutation. First, a comparison of the four germline J_α gene sequences analysed here with their cDNA counterparts from the same inbred strain reveals that there is no somatic mutation (see Fig. 3 and ref. 18). Second, three cDNA clones with identical V_α gene segments have been identified in the 19 different cDNA clones studied and all are found to be identical[18].

The immunogloblulin and T-cell β- and γ-gene families have four distinct types of J-C organization (Fig. 4). (1) The immunoglobulin κ-genes have a cluster of four functional J gene segments located 2.5 kb to the 5′ side of a single C_κ gene[27,28]; (2) the H-chain genes have four J_H gene segments lying 6.5 kb 5′ to the C_μ gene, the most 5′ of eight different C_H genes[29,30]; (3) the λ family has four single J gene segments, each lying 1.2–1.3 kb 5′ of separate C genes[31]. Similarly, the γ-gene family has three single J gene segments, each lying 3.8 kb 5′ of separate C genes[5]; (4) the β gene family has two clusters of J gene segments, each with six functional members. These clusters are located ~2.5 kb to the 5′ side of the two C_β gene segments[7,10,11]. In each case, when multiple J gene segments exist, they are closely clustered, separated by 36–519 bp, with most of the J gene segments separated by <350 bp (Fig. 4). We have demonstrated here that the 18 J_α sequences map across 60 kb of DNA. As 18 out of 19 J_α gene segments analysed are different, many additional J_α gene segments will probably be found in this region. This is in striking contrast to the other five rearranging immunoglobin families where the J gene segments are relatively close to the C gene (Fig. 4).

The disperse organization of the J_α gene segments has several consequences. (1) The size of nuclear transcripts for immunoglobulin and T-cell receptor genes is determined by the length of the J-C intron. In the immunoglobulin, β- and γ-gene families, these introns vary between 1.2 and 8 kilobases. In contrast, the homologous introns of the α-gene vary between 3 and >63 kb. Thus, the α-transcripts must vary from 6 to >66 kb. Several other eukaryotic genes have larger nuclear transcripts[32–34]. (2) The large distance between the rearranged V_α gene and the C_α gene may have interesting regulatory implications. In immunoglobulin H-chain and κ light-chain genes, enhancer sequences are found in the intron between the 3′-most J and the C sequences[35–37]. If similar enhancer sequences exist in the α-gene family, the enhancer sequence could be ≥60 kb from the promoter on which it presumably operates. This possibility raises interesting questions about the distances over which enhancers can operate. Alternatively, there may be multiple enhancer sequences, each associated with a J_α gene segment. (3) Three J_α gene segments, J_α37, J_α57 and J_α61, are closely linked and seem to be most closely related to one another (see ref. 18). This observation suggests that the J_α gene segments fall into subfamilies, like the V_α gene segments. Hence, the duplication, expansion and diversification of J_α gene segments may also arise from unequal crossing-over, gene conversion and single base substitutions. (4) As T cells may not use somatic hypermutation[24], there seem to be compensatory mechanisms for increased diversification of the 3′ ends of the V_β and V_α genes. In the β-genes this occurs in the increased repertoire of J_β gene segments and the extensive use of the D_β gene segments in all three translational reading frames[24]. In the α-genes this occurs in the extensive J_α repertoire.

We thank Marilyn Tomich and Dr Suzanna Horvath for the synthetic oligonucleotides, Henry Sun for the B10.D2-H-2^{dm1} cosmid library, Debbie Maloney and Marty Garcia for technical assistance, and Gerald Siu, Drs Richard Barth, Sandra Chang, Joan Kobori, Mitchell Kronenberg, Ulf Landegren, Nilabh Shastri and Martha Zuniga for critical reading of the manuscript. We also thank Dr Daniel Hansburg for providing the cytochrome $c/1$-$E^\kappa_\alpha E^b_\beta$-specific 1.9.2 T-helper hybridoma for the construction of cDNA and genomic libraries, Gerald Siu for stimulating discussions and Bertha Jones, Connie Katz and Susan Mangrum for preparation of the manuscript. This work was supported by a grant from the NIH and by T Cell Sciences, Inc.

In the accompanying papers, Hayday *et al.*[47] and Yoshikai *et al.*[48] have made similar observations concerning the organization of α genes in mice and humans, respectively.

Received 10 May; accepted 27 June 1985.

1. Acuto, O. *et al. Cell* **34**, 717–726 (1983).
2. Kappler, J., Kubo, R., Haskins, K., White, J. & Marrack, P. *Cell* **34**, 727–737 (1983).
3. McIntyre, B. W. & Allison, J. P. *Cell* **34**, 739–746 (1983).
4. Saito, H. *et al. Nature* **309**, 757–762 (1984).
5. Hayday, A. C. *et al. Cell* **40**, 259–269 (1985).
6. Kranz, D. M. *et al. Nature* **313**, 752–755 (1985).
7. Malissen, M. *et al. Cell* **37**, 1101–1110 (1984).
8. Siu, G. *et al. Cell* **37**, 393–401 (1984).
9. Siu, G. *et al. Nature* **311**, 344–350 (1984).
10. Chien, Y., Gascoigne, N. R. J., Kavaler, J., Lee, N. E. & Davis, M. M. *Nature* **309**, 322–326 (1984).
11. Gascoigne, N. R. J., Chien, Y., Becker, D. M., Kavaler, J. & Davis, M. M. *Nature* **310**, 387–391 (1984).
12. Hedrick, S. M., Cohen, D. I., Nielsen, E. A. & Davis, M. M. *Nature* **308**, 149–153 (1984).
13. Hedrick, S. M., Nielsen, E. A., Kavaler, J., Cohen, D. I. & Davis, M. M. *Nature* **308**, 153–158 (1984).
14. Yanagi, Y. *et al. Nature* **308**, 145–149 (1984).
15. Chien, Y. *et al. Nature* **312**, 31–35 (1984).
16. Saito, H. *et al. Nature* **312**, 36–40 (1984).
17. Sim, G. K., *et al. Nature* **312**, 771–775 (1984).
18. Arden, B., Klotz, J., Siu, G. & Hood, L. *Nature* **316**, 783–787 (1985).
19. Early, P., Huang, H., Davis, M. M., Calame, K. & Hood, L. *Cell* **19**, 981–992 (1980).

20. Strauss, E. C., Kobori, J. A., Siu, G. & Hood, L. *Analyt. Biochem.* (submitted).
21. Breathnach, R. & Chambon, P. *A. Rev. Biochem.* **50**, 349-383 (1981).
22. Caccia, N. *et al. Cell* **37**, 1091-1099 (1984).
23. Alt, F. W. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4118-4122 (1982).
24. Barth, R. K. *et al. Nature* **316**, 517-523 (1985).
25. Kim, S., Davis, M., Sinn, E., Patten, P. & Hood, L. *Cell* **27**, 573-581 (1981).
26. Gearhart, P. J., Johnson, N. D., Douglas, R. & Hood, L. *Nature* **291**, 29-34 (1981).
27. Sakano, H., Huppi, K., Heinrich, G. & Tonegawa, S. *Nature* **280**, 288-294 (1979).
28. Max, E. E., Maizel, J. V. & Leder, P. *J. biol. Chem.* **256**, 5116-5120 (1981).
29. Sakano, H., Maki, R., Kurosawa, Y., Roder, W. & Tonegawa, S. *Nature* **286**, 676-684 (1980).
30. Newell, N., Richards, J. E., Tucker, P. W. & Blattner, F. R. *Science* **209**, 1128-1132 (1980).
31. Blomberg, B. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **79**, 530-533 (1982).
32. Bender, W. *et al. Science* **221**, 23-29 (1983).
33. Gitschier, J. *et al. Nature* **312**, 326-330 (1984).
34. Scott, M. P. *et al. Cell* **35**, 763-776 (1983).
35. Gillies, S. D., Morrison, S. L., Oi, V. T. & Tonegawa, S. *Cell* **33**, 717-728 (1983).
36. Banerji, J., Olson, L. & Schaffner, W. *Cell* **33**, 729-740 (1983).
37. Queen, C. & Baltimore, D. *Cell* **33**, 741-748 (1983).
38. Huynh, T. V., Young, R. A. & Davis, R. W. in *DNA Cloning: A Practical Approach* (ed. Glover, D.) (IRL, Oxford, 1984).
39. Okayama, H. & Berg, P. *Molec. cell. Biol.* **2**, 161-170 (1982).
40. Steinmetz, M., Winoto, A., Minard, K. & Hood, L. *Cell* **28**, 489-498 (1982).
41. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning* (Cold Spring Harbor Laboratory, New York, 1982).
42. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
43. Frischauf, A., Lehrach, H., Poustka, A. & Murray, N. *J. molec. Biol.* **170**, 827-842 (1983).
44. Yanisch-Perron, C., Vieira, J. & Messing, J. *Gene* **33**, 103-119 (1985).
45. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1980).
46. Shimizu, A., Takahashi, N., Yaoita, Y. & Honjo, T. *Cell* **28**, 499-506 (1982).
47. Hayday, A. C. *et al. Nature* **316**, 828-832 (1985).
48. Yoshikai, Y. *et al. Nature* **316**, 837-840 (1985).

Chapter Five

ANALYSIS OF MOUSE T_H CELL RECEPTORS SPECIFIC FOR PIGEON CYTOCHROME C AND RESTRICTED BY THE MHC E MOLECULE: PREDOMINANT USAGE OF A V_α GENE SEGMENT

Submitted to *Cell*

Analysis of Mouse T_H Cell Receptors Specific for Pigeon Cytochrome c and Restricted by the MHC E Molecule: Predominant Usage of a V_α Gene Segment

Astar Winoto, Nancy C. Lan, James L. Urban, Joan Goverman, Leroy Hood, and Daniel Hansburg*

Division of Biology

California Institute of Technology

Pasadena, California 91125

*Department of Pathology

Fox Chase Cancer Center

7701 Burholme Ave.

Philadelphia, Pennsylvania 19111

Running title: Cytochrome c/I-E-specific T-cell receptor genes

Summary

We have determined the nucleotide sequences of seven functional α and six functional β chains of the T-cell receptor genes from nine T-helper hybridomas specific for the C-terminal peptide of pigeon cytochrome c. Northern blot analyses using the isolated V_α and V_β gene segments were performed on the RNAs isolated from a total of 15 T-helper hybridomas with a similar specificity. A single V_α subfamily is predominantly used in these 15 T-helper hybridomas, whereas at least five different V_β gene segments are utilized. The patterns of T-cell gene segment usage and alterations in antigen specificity associated with changes in the third hypervariable region are similar to those seen in B-cell immunoglobulin responses to antigen, consistent with the idea that the T-cell receptor and immunoglobulin molecules are similar in structure to one another. However, in contrast to B-cell receptor diversification, somatic hypermutation does not appear to play a crucial role in the diversification of T-cell receptor α or β chains specific for cytochrome c.

Introduction

The vertebrate immune system employs two classes of clonally differentiated, antigen-specific lymphocytes, B cells and T cells. The antigen receptors of each class of cells, immunoglobulins and the T-cell receptors, respectively, are glycoproteins and share many properties. Both are heterodimers (heavy/light and α/β) with individual chains divided into variable and constant regions. Each variable region is encoded by either two or three gene segments: a variable (V) gene segment, a joining (J) gene segment and for the heavy and β chains, a diversity (D) gene segment. These gene segments rearrange and join during lymphocyte differentiation to form a functional V gene (VJ or VDJ). The V gene segments fall into distinct subfamilies with the members of a particular subfamily at least 75% homologous at the DNA level. The 3' ends of the V_α and V_β genes are very diverse. In the mouse, there are probably more than 50 J_α and 12 J_β gene segments (Chien et al., 1984; Gascoigne et al., 1984; Malissen et al., 1984; Arden et al., 1985; Hayday et al., 1985; Winoto et al., 1985; Yoshikai et al., 1985), whereas there are only four J_L and four J_H gene segments. Both the T-cell and B-cell receptors are diversified by a variety of mechanisms: multiplicity of germline gene segments, combinatorial joining of gene segments, the random addition of nucleotides between the rearranging gene segments (N-region diversification), and variation in the sites at which the gene segments are joined (junctional variation) (for review see Tonegawa, 1983 and Kronenberg et al., 1986). One striking difference between the V_β genes and their immunoglobulin counterparts is that somatic hypermutation which can alter up to 3% of a given V gene sequence (Kim et al., 1981; Pech et al., 1981) is not generally seen in the T-cell receptor V_β genes (Chien et al., 1984; Barth et al., 1985; Behlke et al., 1985). Preliminary data suggest that V_α genes may also fail to exhibit somatic hypermutation (Hayday et al., 1985; Winoto et al., 1985).

T-cell antigen recognition differs from B-cell antigen recognition in that the T cell recognizes antigen in the context of cell surface molecules encoded by the major histocompatibility complex (MHC), a phenomenon termed MHC restriction (Kindred and Shreffler, 1972; Katz et al., 1973; Rosenthal and Shevach, 1973; Zinkernagel and Doherty, 1974). T-helper cells (T_H), typically restricted to the class II MHC (I region) gene products (A or E molecules), function to support the differentiation and proliferation of B cells and cytotoxic T cells in response to antigenic stimulation. The class II A or E molecules are heterodimers composed of highly polymorphic α and β chains. The constellation of MHC alleles or haplotypes of particular strains of mice are designated as superscript letters above each chain; e.g., inbred B10.A mice expressing E molecules of the k haplotype are denoted as $E_{\alpha}^{k}E_{\beta}^{k}$.

A recent experiment has demonstrated that the transfer of particular α and β genes into a T-cell hybridoma can reconstitute the donor T-cell specificity for antigen and the MHC molecule (Dembic et al., 1986), thus suggesting that the α/β complex alone controls T-cell specificity. Several analyses of the T-cell receptor V regions suggest that their secondary structures are similar to those of the immunoglobulins. First, immunoglobulin V_L and V_H regions have 14 highly conserved amino acids believed to play an important role in intra- and interchain interactions. These same residues are highly conserved in the V_{α} and V_{β} regions (for review, see Kronenberg et al., 1986). Second, the potential for forming β-pleated sheets and hydrophobicity patterns are similar in the V_L ,V_H, V_{α} and V_{β} regions (Patten et al., 1984; Arden et al., 1985; Barth et al., 1985; Becker et al., 1985; Behlke et al., 1985). These observations suggest that the three-dimensional structure of the V domains of the T-cell receptors are similar to those of their immunoglobulin counterparts. Moreover, the V_{β} region appears to have hypervariable regions at positions similar to those of immunoglobulins, and the V_{α}

region has at least one hypervariable region analogous to the immunoglobulin third hypervariable region. These hypervariable regions presumably fold to constitute the walls of the antigen/MHC-binding cleft (Patten et al., 1984; Arden et al., 1985; Barth et al., 1985; Becker et al., 1985; Behlke et al., 1985). Accordingly, the T-cell receptor molecule probably bears an antigen-binding pocket similar to that of antibodies, allowing it to recognize an interaction surface formed by the juxtaposition of an antigen and a MHC molecule (Schwartz, 1985; Goverman et al., submitted).

To assess the relative contribution of the V_α and V_β genes of the mouse T-cell receptor to the recognition of a limited antigenic determinant and two different MHC molecules, we have cloned and sequenced seven functional V_α and six functional V_β genes from nine T-helper hybridomas specific for the C-terminal peptide of the pigeon cytochrome c. Northern blot analyses were performed on the RNAs from a total of 15 T-helper hybridomas specific for the cytochrome c peptides using the isolated V_α and V_β gene segments as probes. A single V_α gene segment is predominantly used in the T-cell receptors of these T-cell hybridomas in conjunction with five different V_β gene segments. The $V_\beta D_\beta J_\beta$ and probably the $V_\alpha J_\alpha$ junctional sequences are important in altering the antigen fine specificities of the T-cell clones, suggesting that the third hypervariable region forms a portion of the antigen-binding cleft. Finally, six independently derived V_α gene segment sequences are identical, suggesting that somatic hypermutation does not operate in the V_α gene of the T-cell receptor.

Results and Discussion

Strategy

The immunoglobulin responses to several different antigens, both haptens and proteins, have been characterized (Table 1). The immune responses to haptens

generally employ a single V_H or V_L gene segment, although the oxazolane response is more complex. The response to the haemaglutinun protein determinant employs a single V_L gene segment and three V_H, three J_H and three J_L gene segments. We were interested in determining the patterns of gene segment usage for T cells specific for a protein determinant and to what extent they correlated with antigen or MHC specificity.

We chose to examine the diversity of the V_α and V_β genes in the T-cell receptors of T_H cell specific for pigeon cytochrome c for several reasons (for review, see Schwartz, 1985). First, T_H cells specific for the C-terminal peptide of cytochrome c in the context of an appropriate E molecule may be stimulated to produce the hormone IL-2 (Matis et al., 1982). This is a simple and sensitive assay system. Second, the C-terminal peptide, residues 81-104 of pigeon cytochrome c contain the dominant antigenic determinant in the T-cell responses to this antigen (Solinger et al., 1979; Hansburg et al., 1983). Accordingly, most T cells interact with a limited determinant on this protein molecule. This specificity is supported by the observation that a change of lysine 99 to arginine, glutamine, or the acetimidyl form of lysine changes the responding T_H specificity (Hansburg et al., 1981, 1983). Third, inbred mice of the k (B10.A) or v (B10.SM) haplotypes respond to this antigen. Accordingly, the diversity of T-cell receptors binding to the same antigen in the context of different MHC molecules can be assessed. Fourth, the T-cell response to cytochrome c is degenerate for certain MHC molecules presenting the antigen. The T-cell clones generated from B10.A or B10.A(5R) mice can be stimulated with the cytochrome c peptide in conjunction with either the $E^k_\alpha E^k_\beta$ molecule encoded by B10.A mice or the $E^k_\alpha E^b_\beta$ molecule encoded by B10.A(5R) mice (Heber-Katz et al., 1982). The T-cell clones produced from B10.S(9R) or B10.SM mice can be stimulated with the cytochrome c peptide in conjunction with either the $E^k_\alpha E^s_\beta$ molecule encoded by the B10.S(9R) mice or the

$E^V_\alpha E^V_\beta$ molecule encoded by the B10.SM mice. This degeneracy of MHC restriction presumably occurs because of the extensive similarities among certain pairs of E_β alleles (Mengle-Gaw et al., 1984, 1985). Because of these cross-reactivities and for simplicity, we will term the T-cell hybridomas generated from either B10.A or B10.A(5R) mice as restricted to the E^k molecule and the T-cell hybridomas produced from either B10.S(9R) or B10.SM mice as restricted to the E^s molecule. Fifth, the subtle changes in specificity of differing T-cell receptors can be probed using a panel of peptides that are closely related to the pigeon C-terminal peptide (Figure 1). Studies with these peptides have demonstrated that residue 100 (Solinger et al., 1979) as well as residues 99 and 103 (Hansburg et al., 1983) are critical for the mouse T_H-immune response to pigeon cytochrome c. Accordingly, sequence changes in the V_α or V_β regions may be reflected by changes in fine specificity of the T-cell clones to this panel of peptides.

The T_H hybridomas analyzed in this report were generated by five different cell fusions carried out in four different strains of mice immunized with pigeon cytochrome c or with the C-terminal homologous peptides denoted DASp or Sp (Table 2 and Figure 1). Hence, the T-cell hybridomas we analyzed can be stimulated to produce IL-2 when presented with the C-terminal peptide containing lysine 99 in the context of the E^k molecule for eight of the hybridomas and in the context of the E^s MHC molecule for the remaining six.

Isolation and Analysis of the T-cell Receptor V_α Genes

The functional V_α genes were isolated from seven T_H hybridomas using three different strategies. First, cDNA libraries constructed from hybridomas V1.9.2, V11.5 and AP11.2 in the vector λgt10 were screened with a C_α probe. Second, the V_α rearrangements in hybridomas AN6.2 and BC15.1 were detected by Southern blot analyses with several restriction fragments containing J_α gene segments as probes (Winoto et al., 1985). The rearranged bands were then cloned into the

vector λEMBL3 and screened with the J_α probes. Third, Northern blot analyses of the RNAs from hybridomas AN14.4 and BC37.5 demonstrated that the AN6.2 V_α gene segment was expressed (Figure 2). The rearranged bands in these hybridomas were identified by a 5' flanking probe from the AN6.2 V_α gene (Figure 3). The rearranged bands were then cloned in the vector λL47.1 and screened with the flanking probe from the AN6.2 V_α gene.

The V_α genes obtained from seven hybridomas were sequenced by the dideoxy chain termination method. Six of the V_α gene segments were identical in sequence, differing from the previously published V_α gene segment expressed by the cytochrome c/I-E^k-specific hybridoma 2B4 by nine nucleotides in the coding region and nine nucleotides in the 5' flanking region (Becker et al., 1985 and Figure 4A). Since hybridoma 2B4 also originated from a C57BL/10 background, the differences in the V_α gene sequences cannot be attributed to strain polymorphism. Hence the two V_α gene segments are distinct members belonging to the same V_α subfamily, here designated as $V_{\alpha 11.1}$ and $V_{\alpha 11.2}$. The remaining hybridoma (AP11.2) employed a V_α gene segment belonging to the $V_{\alpha 4}$ subfamily which differs from the $V_{\alpha 11}$ gene segment of hybridoma AN6.2 by 80% at the protein level and by 61% at the DNA level (Arden et al., 1985; Figure 4B). The $V_{\alpha 11}$ subfamily contains five members and the $V_{\alpha 4}$ subfamily contains at least nine members (Arden et al., 1985; Becker et al., 1985 and Winoto, unpublished data). In contrast, for the seven hybridomas using receptors with $V_{\alpha 11}$ gene segments, there were five different J_α gene segments expressed; one ($J_{\alpha 84}$) was used three times; the other four were used only once. The hybridoma AP11.2, which uses a receptor containing a $V_{\alpha 4}$ gene segment, utilizes yet another J_α gene segment (Table 3 and Figure 4).

Isolation and Analysis of V_β Genes

To isolate the functional V_β genes, Southern genomic blot analyses of the

hybridoma DNAs were performed using four restriction fragments as probes: $J_{\beta 1}$, $J_{\beta 2}$, $D_{\beta 1}$ and $D_{\beta 2}$ probes. The $J_{\beta 1}$ and $J_{\beta 2}$ restriction fragments containing $J_{\beta 1.3\text{-}1.6}$ and $J_{\beta 2.1\text{-}2.6}$ gene segments, respectively, detect all the rearranged β genes. The D_β probes consist of DNA fragment 5' to the D_β gene segments and detect only the partial $D_\beta J_\beta$ rearrangements. The rearranged bands that hybridize with the $J_{\beta 1}$ or $J_{\beta 2}$ probes and do not hybridize with the D_β probes presumably contain the $V_\beta D_\beta J_\beta$ rearranged genes. The bands were subsequently isolated and cloned in the λgt7 lac5, λL47.1, or λ590 vectors. The libraries were then screened with the appropiate J_β probe.

The isolated clones were then sequenced using the oligonucleotides specific for the corresponding V_β and J_β gene segments. The V_β genes from eight cytochrome c-specific hybridomas, including two published V_β genes from hybridomas V1.9.2 and 2B4 (Chien et al., 1984; Barth et al., 1985) were compared (Table 3 and Figure 5). Among these eight hybridomas, three different V_β gene segments ($V_{\beta 1}$, $V_{\beta 3}$ and $V_{\beta 8.3}$) and four different J_β gene segments ($J_{\beta 1.1}$, 1.2, 1.4 and 2.5) were used. The $V_{\beta 3}$ and $J_{\beta 1.2}$ gene segments were used more frequently than the others (five times).

The $V_{\alpha 11}$ Gene Segment Subfamily is Predominantly Employed in T_H Responses to the C-terminal Peptide of Cytochrome c

The V_α gene segments from hybridomas V1.9.2 (subfamily $V_{\alpha 11}$) and AP11.2 (subfamily $V_{\alpha 4}$) were used as probes in Northern blot analyses of seven additional T_H hybridomas specific for the C-terminal peptide of cytochrome c. Two of these (V15.4 and AN4.4) are restricted to the E^k molecule and the remainder to the E^s molecule. One of the I-E^k-restricted and three of the I-E^s-restricted T cells employ the $V_{\alpha 11}$ gene segment, whereas the remainder do not employ either $V_{\alpha 11}$ or $V_{\alpha 4}$ gene segments (Figure 2 and Winoto, unpublished observation). Moreover, Southern genomic blot analysis using a $V_{\alpha 11}$ flanking region probe suggests that at

least one of the I-E^S restricted hybridomas (AR8.1) uses the $V_{\alpha 11.1}$ gene segment (Figure 3). Thus, eight out of nine cytochrome c-specific T_H hybridomas restricted to the E^k molecule and three out of six T_H hybridomas restricted to the E^S molecule express a gene segment of the $V_{\alpha 11}$ subfamily (Table 4). At least seven of them employ the $V_{\alpha 11.1}$ gene segment. Accordingly, the cytochrome c responses studied here employ predominantly T-cell receptors containing α chains encoded by the $V_{\alpha 11}$ subfamily.

Lack of Absolute Correlation of α or β Chain Gene Segments to Antigen Recognition or MHC Restriction

To survey the distribution of V_β gene segment usage in the cytochrome c-specific T-helper hybridomas, Northern blot analyses of the RNAs from a total of 18 T_H hybridomas using probes from the $V_{\beta 2, 3, 4, 5, 8 \text{ and } 14}$ subfamilies (nomenclature as in Barth et al., 1985; Malissen et al., 1986) were carried out (Table 4). The $V_{\beta 1}$ probe was not used because the fusion partner BW5147 uses this gene segment. These data were then combined with those from our sequence analyses of V_β gene segments. Out of 12 I-E^k restricted T-helper hybridomas, nine of them were specific for the DASp peptide and the remaining three were specific for the DAQ^{99} peptide (see Fig. 1). At least five different V_β gene segments were employed in the nine DASp/I-E^k-specific hybridomas ($V_{\beta 1}$, $V_{\beta 3}$, $V_{\beta 8}$, $V_{\beta 14}$ and x, where x denotes unassigned subfamilies) and three different V_β gene segments were used in the three DAQ^{99}/I-E^k-specific hybridomas ($V_{\beta 2}$, $V_{\beta 8}$ and x). Of the six DASp/I-E^S-specific T_H hybridomas, at least four different V_β segments were employed ($V_{\beta 1}$, $V_{\beta 3}$, $V_{\beta 8}$ and x). Thus, the same V_β gene segments may be employed in binding 1) a similar antigen (DASp) in the context of two different MHC molecules (E^k and E^S) or 2) different antigens (DASp and DAQ^{99}) in the context of the same MHC molecule (E^k, see Table 4). Likewise, there is no correlation of J_β gene segments with antigen specificity or MHC restriction

(Tables 3 and 4). Although there is a predominate usage of the $V_{\alpha 11.1}$ gene segment in the binding of the same antigen (DASp) with two different MHC molecules, at least four different V_α gene segments are employed in this response ($V_{\alpha 11.1}$, $V_{\alpha 11.2}$, $V_{\alpha 4.3}$ and x). Hence there is no absolute correlation of the V_α gene segment to either antigen or MHC specificity. Insufficient J_α sequences have been determined to rule out a correlation with the MHC restricting elements, however, the diversity of the J_α gene segments employed in the I-E-restricted response once again suggests there will be no simple correlation (Table 3).

Correlation of T-cell Receptor Structure with the Antigen Fine Specificity

Several lines of evidence suggest that the V_α region may interact with the epitope containing lysine 99 of the cytochrome c peptide. First, three out of three hybridomas generated in B10.A mice with glutamine at position 99 (DAQ^{99} peptide, Figure 1) do not employ a $V_{\alpha 11}$ gene segment (Winoto, unpublished observation), whereas 11/15 hybridomas raised against antigens with a lysine at position 99 do use $V_{\alpha 11}$ gene segments. Second, when the lysine at position 99 is altered in the peptide presented to these 15 T_H cells, 14 fail to be stimulated (hybridomas V1.9.2 is an exception, see Table 2).

Other data indicate that changes in the third hypervariable regions may lead to changes in antigenic specificity. (The third hypervariable region for the α chain includes the $V_\alpha J_\alpha$ junctional region and the N-terminal portion of the J_α region. For the β chain it includes the $V_\beta D_\beta J_\beta$ junctional region, all of D_β and the N-terminal of the J_β segments.) The V_α and V_β chains from hybridomas AN6.2 and V11.5 are identical but for five amino acid differences in the third hypervariable region of the V_β chain (Fig. 5B). To determine how these T_H hybridomas differed in their antigenic fine specificity, six different related cytochrome c peptides (Figure 1) were used to compare the reactivities of hybridomas V11.5 and AN6.2

(Figure 6). The specificity was determined by measuring the relative concentrations of different antigens needed for identical IL-2 production by the T-cell hybridomas as measured by the uptake of ^{3}H-thymidine by a growing IL-2-dependent cell line. The most notable difference was obtained for peptide DAK^{100}, a derivative of the DASp peptide in which the lysine at position 100 was changed to glutamine. Hybridoma V11.5 did not respond to the DAK^{100} peptide even at high doses of antigen concentration, while hybridoma AN6.2 reacted strongly with this peptide in comparison to the DASp and pigeon cytochrome c response. The Sp peptide, an alanine^{103} analog of the DASp peptide, revealed additional differences between the specificities of hybridomas V11.5 and AN6.2. Hybridoma V11.5 reacted to the DASp peptide more effectively than to the Sp peptide, whereas hybridoma AN6.2 reacted equally to both peptides. Since both T cells are restricted to the E^{k} molecule we conclude that changes limited to the third hypervariable region of the V_{β} sequence may alter the antigen fine specificity of individual T-cells without altering MHC specificity.

The T-cell receptors of hybridomas AN6.2 and AN14.4 are identical but for differences in the third hypervariable regions of both their V_{α} (ten differences) and V_{β} (two substitutions) regions. The antigen fine specificities of hybridomas AN6.2 and AN14.4 were then compared (Figures 6E and 6B). The AN6.2 hybridoma reacted strongly to the DAK^{100} peptide and hybridoma AN14.4 reacted weakly only at high concentration. Moreover, the AN14.4 hybridoma responded to the DASp peptide more effectively than the Sp peptide, whereas the AN6.2 hybridoma responded equally to the two peptides. It is obviously impossible to determine whether one or both hypervariable regions account for these differences in antigenic specificity. However, once again both hybridomas are restricted to the E^{k} MHC molecule, so these differences alter antigenic but not MHC specificity.

The important conclusion to be drawn from these studies is that changes in the third hypervariable regions of the V_β and possibly the V_α regions lead to differences in antigen fine specificities, just as with immunoglobulins. Once again, these data are consistent with the hypothesis that immunoglobulin and T-cell receptors fold into similar three-dimensional configurations and interact with antigen (antigen MHC) through their hypervariable regions.

Different V_α and V_β Genes May Generate Similar Fine Specificities

Since hybridoma AN4.4 is the only T-helper hybridoma restricted to the E^k molecule that does not use a $V_{\alpha 11}$ gene segment, we were interested in determining whether this hybridoma exhibits unique antigen fine specificity. Fine specificity studies were performed on five hybridomas using six different cytochrome c peptides (Figure 1) with B10.A(I-E^k) antigen-presenting cells. The results from hybridoma AN4.4 (Figure 6A) were compared to four other $V_{\alpha 11}$-expressing hybridomas from three different fusions. We found that one of the $V_{\alpha 11}$-expressing hybridomas, AN14.4 (Figure 6B), has a similar profile of antigen specificities to hybridoma AN4.4. The DAK^{100}, a (Gln to Lys)^{100} substituted analog of the DASp peptide, was a moderately stronger antigen for hybridomas AN4.4 and AN14.4, a very strong antigen for hybridoma AN6.2, and a weak antigen for the other two hybridomas BC37.5 and V11.5 (Figures 6C and 6D). The fly cytochrome c C-terminal peptide, containing Ser^{100}, was a very strong antigen for hybridomas BC37.5 and V11.5 but not for the AN hybridomas. Hence, different V_α segments in association with different V_β segments in hybridomas AN4.4 and AN14.4 can give rise to similar cytochrome c fine specificities.

A similar conclusion was reached when we examined the differences between the T-cell receptors of pairs of hybridomas with fine specificity patterns as closely matched as possible. The fine specificity of hybridoma V11.5 has been discussed previously and is identical on B10.A (I-E^k) antigen presenting cells by

the currently available peptide panel to that of hybridoma BC37.5 (Figures 6D and 6C). Hybridomas AP11.2 and V15.4 also exhibit similar fine specificity, although AP11.2 is restricted to the E^s molecule and hybridoma V15.4 is restricted to the E^k molecule (Figures 7A and 7B). Obviously, the antigenic fine specificity analysis is limited by the size of the panel of peptides tested. Nevertheless, it is remarkable that these pairs of T_H hybridomas are so similar in their fine specificity, and yet sequencing, and Northern blot analyses, indicate that the two pairs of hybridomas do not share the same pairs of V_α/V_β gene segments. This suggests that a combination of different V_α V_β regions can form similar antigen/MHC-binding pockets.

Somatic Hypermutation Does Not Operate In the T-cell Receptor Genes

Analyses of the six V_α gene segments used by the T-helper hybridomas specific for the C-terminal peptide of cytochrome c do not show any nucleotide changes (Figure 4). All six T cells are of independent origin from three different inbred mice with the C67BL/10 T-cell receptor genes. Since all six T-helper hybridomas are functional and presumably represent mature T cells, we conclude that somatic hypermutation occurs infrequently, if at all, in the α genes of T_H cells specific for cytochrome c. Similar conclusions have been reached for the β genes of the T-cell receptor (Chien et al., 1984; Barth et al., 1985; Behlke et al., 1985). Our data are consistent with this latter interpretation in that four $V_{\beta 3}$ gene segments used by four different T-helper hybridomas specific for cytochrome c are identical in their coding region nucleotide sequences except for one mutation at position 95 of the $V_{11.5}$ V_β gene segment (Figure 5). Somatic hypermutation in B cells appears to occur relatively late in B-cell differentiation. Our T-cell hybridomas were generated from T cells that were stimulated once *in vivo* and a week later placed in tissue culture with antigen for three days prior to fusion. Although it is possible that somatic hypermutation has not occurred due to the fact that our T

cells are analogous to primary B cells or due to the antigen excess condition *in vitro*, the data here and data of 27 human V_β sequences from peripheral blood (Concannon et al., manuscript in preparation) are consistent with the supposition that somatic hypermutation does not play an important role in generating the diversity of the T-cell receptor repertoire.

Pattern of the Cytochrome c Specific T-cell Receptors V Gene Usage Resembles that of Their Immunoglobulin Counterparts

Data from study of the V_α, V_β gene usage in the T_H responses to cytochrome c are consistent with the pattern of gene segment usage seen in immunoglobulins and exclude a simple dual site single receptor model of T-cell recognition (e.g., the α gene segments correlates with MHC restriction and the β chain gene segments with the antigen or vice versa). Fifteen T cells from several strains of mice immunized with pigeon cytochrome c or the closely related terminal peptides, appear to employ a predominant V_α, five or more V_β, six or more J_α and at least four different J_β gene segments. The predominant usage of one V gene segment is similar to that seen in some specific antibody responses (Table 1). Hence, there are three possible patterns of the V gene segment usage in other T-cell responses. First, most specific T-cell responses may employ a predominant V_α gene segment, suggesting that the α chain may contribute more to determine the complementary site for antigen binding. Alternatively, other T-cell responses may show a predominant V_β gene segment or no predominance at all of either α or β chains. These are the patterns seen in the specific antibody responses.

These data raise interesting clinical possibilities. Perhaps autoimmune diseases, where T-cell reactivity against a specific cellular antigen has been implicated, will be associated with predominant usage of a particular T-cell receptor V_α (or V_β) gene segment. Antibodies to these V gene segments might then have useful diagnostic or therapeutic possibilities.

Experimental Procedures

Construction of cDNA and Genomic Library

The cDNA libraries were constructed in the λgt10 vector as described (Winoto et al., 1985). The genomic limited libraries were constructed in λgt7 lac5, λL47.1, λ590 or EMBL3 vectors. Briefly, the arms of λgt7 lac5 or λ590 were separated on low melting agarose gel and isolated by several phenol and chloroform extractions followed by ethanol precipitation. The Bam HI or Hind III arms of λL47.1 were prepared as follows: λL47.1 DNA were digested with Bam HI or Hind III enzyme in combination with Xho I and Sal I enzymes. The digested DNA was then phenol extracted and ammonium/isopropanol precipitated four times at room temperature to remove the small Sal I/Xho I fragment. The DNA was then resuspended in water. λ EMBL3 arms were prepared similarly to that of λL47.1, except that the DNA was cut with Bam HI and Eco RI. The λ arms were ligated to the isolated genomic DNA at a concentration of 250 μg/ml and packaged *in vitro*.

For the isolation of V_α genes from hybridomas AN6.2 and BC15.1, the 15 kb Bam HI rearranged bands were cloned into the Bam HI site of the λ EMBL3 vector; and for hybridomas AN14.4 and BC37.5, the 6 kb Bam HI rearranged bands were cloned into the Bam HI site of λL47.1 vector. The V_β genes from hybridomas AN6.2, AN14.4, and AP11.2 were cloned from limited libraries constructed in the λL47.1 vector from 6 kb Bam HI digested, size selected insert DNA of the corresponding hybridoma DNA. The AP15.2 V_β gene was a 4 kb Eco RI fragment cloned from a limited library made in the λgt7 lac5 vector, the V_β gene of hybridoma V15.4 was isolated from an 8 kb Hind III limited genomic library constructed in the λ590 vector, and the V_β gene of hybridoma V11.5 was isolated from the 6 kb Bam HI limited libraries made in λ590.

Southern and Northern Blots and DNA Sequencing

Southern and Northern blots were performed as described (Maniatis et al., 1982).

DNA sequencing was done on both strands using the dideoxy chain termination method with specific oligonucleotide primers (Sanger et al., 1977; Chen and Seeburg, 1985; Strauss et al., 1986).

The following fragments were used for subcloning into M13 mp18 and/or mp19 for sequencing analysis: 1) For V_α genes of hybridoma AN6.2, a 3.4 kb Kpn I/Sal I fragment; hybridoma BC15.1, a 2.6 kb Hind III fragment; hybridoma AN14.4, a 2.1 kb Hind III fragment; and hybridoma BC37.5, a 2.2 kb Hind III fragment. 2) For the V_β genes of hybridomas AN6.2, AN14.4, AP11.2, V15.4, 3.5 kb Hind III/Bam HI fragments and hybridoma AP15.2, a 4 kb Eco RI fragment. The cDNA fragments containing the V_α genes of hybridomas V1.9.2, V11.5 and AP11.2 were subcloned directly into the Eco RI site of the M13 mp8 vector. The V_β gene of hybridoma V15.4 was sequenced in the pUC8 vector as a 8 kb Hind III fragment. The M13 clones containing different strands of DNA were identified by hybridization to either of two complementary oligonucleotides specific for the C_α gene (for the cDNA clones) or the J_β oligonucleotide in conjunction with the nick translated J_β fragment (for the V_β gene containing clones) or two complementary oligonucleotides specific for $J_{\alpha 84}$ gene segment (for the V_α genes of AN6.2 and BC15.1) or two complementary oligonucleotides specific for the $V_{\alpha 11.1}$ gene segment (for the V_α genes of AN14.4 and BC37.5). The cDNA clones were sequenced utilizing the M13 universal primer and a C_α-specific oligonucleotide (CA3) as primer. The genomic V_α and V_β genes were sequenced using four different oligonucleotide primers, three specific for different regions and different strands of the V gene segment and one specific for the J gene segment.

Generation of Hybridomas and IL-2 Assay

Immunization of the mice and the generation of hybridomas were done as described (Hansburg and Appella, 1985). The IL-2 release assay was done using HT-2 cells as described (Hansburg and Appella, 1985).

Some objection could be raised in that the hybridomas we analyzed do not reflect the actual representation of the T cells in the lymph node. This is due to the differences between the specificity patterns of the hybridomas and the specificity patterns of the whole lymph nodes. These differences lie mainly in the stronger response of pigeon cytochrome c rather than the response of fly cytochrome c. There are many potential reasons for these differences. One reason is that the response patterns in short term culture may be dominated by the most sensitive clones (i.e., the highest avidity cells), whereas the results of a survey of hybridomas will be dominated by the most frequently occurring T cells in response to high antigen concentrations. Nonetheless, one of the hybridoma studied (BC15.1) did display a "typical" lymph node type pattern (Figure 7C). This hybridoma displayed a stronger response to the DASp peptide than pigeon cytochrome c at high doses and a stronger response to fly cytochrome c. There was no stimulation with either the DAK^{100}, DAE^{100} or DAQ^{99} peptides. This clone did use the $V_{\alpha 11.1}$ gene segment, and hence our conclusion obtained from the study of T-cell hybridomas probably reflected the actual situation in the lymph nodes.

Acknowledgements

We thank Debbie Maloney and Karyl Minard for excellent technical help, Drs. Richard Barth, Joan Kobori, Mitchell Kronenberg, Ulf Landegren and Nilabh Shastri for the critical reading of the manuscript. We are also grateful to Bertha Jones, Connie Katz and Cathy Elkins for the preparation of this manuscript. We acknowledge the following sources of support: T Cell Sciences, Inc., and NIH (to L.E.H.), NIH grant-in-aid AI-20661 and CA-06927 (to D.H.) and the Leukemia Society of America (to J.G.).

References

Arden, B., Klotz, J., Siu., G., and Hood, L. (1985). Diversity and structure of genes of the α family of mouse T-cell receptor. *Nature 316*, 781-787.

Ball, R. K., Chang, J.-Y., Alkan, S. S., and Braem, D. G. (1983). The complete amino acid sequence of the light chain variable region of two monoclonal anti-p-azobenzene-arsonate antibodies bearing the cross-reactive idiotype. *Mol. Immunol. 20*, 197-201.

Barth, R., Kim, B., Lan, N., Hunkapiller, T., Sobieck, N., Winoto, A., Gershenfeld, H., Okada, C., Hansburg, D., Weissman, I., and Hood, L. (1985). The murine T-cell receptor employs a limited repertoire of expressed Vβ gene segments. *Nature 316*, 517-523.

Becker, D., Patten, P., Chien, Y.-H., Yokoita, T., Eshbar, Z., Giedlin, M., Gascoigne, N. R. J., Goodenow, C., Wolf, R., Arai, K.-I., and Davis, M. M. (1985). Variability and repertoire size in T-cell receptor Vα gene segments. *Nature 317*, 430-434.

Behlke, M. A., Spinella, D. G., Chou, H., Sha, W., Hartl, D. L., and Loh, D. Y. (1985). T-cell receptor β chain expression dependence on relative few variable genes. *Science 229*, 566-570.

Berek, C., Griffiths, G. M., and Milstein, C. (1985). Molecular events during maturation of the immune response to oxazolone. *Nature 316*, 412-418.

Chen, E. Y., and Seeburg, P. H. (1985). Supercoiling sequencing: A fast and simple method for sequencing plasmid DNA. *DNA 4*, 165-170.

Chien, Y.-H., Gascoigne, N. R. J., Kavaler, J., Lee, N. E., and Davis, M. M. (1984). Somatic recombination in a murine T-cell receptor gene. *Nature 309*, 322-326.

Clarke, S. H., Huppi, K., Ruezinsky, D., Staudet, L., Gerhard, W., and Weigert, M. (1985). Inter- and intraclonal diversity in the antibody response to influenza haemaglutinin. *J. Exp. Med. 161*, 687-704.

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

Gascoigne, N. R. J., Chien, Y.-H., Becker, D. M., Kavaler, J., and Davis, M. M. (1984). Genomic organization and sequence of T-cell receptor β -chain constant- and joining-region genes. *Nature 310*, 387-391.

Hansburg, D., Hannum, C., Inman, J. K., Appella, E., Margoliash, E., and Schwartz, R. H. (1981). Parallel cross-reactivity patterns of two sets of antigenically distinct cytochrome c peptides: Possible evidence for a presentational model of Ir gene function. *J. Immunol. 127*, 1844-1851.

Hansburg, D., Fairwell, T., Schwartz, R. H., and Appella, E. (1983). The T-lymphocyte response to cytochrome c. IV. Distinguishable sites on a peptide antigen which affect antigen strength and memory. *J. Immunol. 131*, 319-324.

Hansburg, D., and Appella, E. (1985). The sites of antigen-T cell and antigen-MHC interactions overlap. *J. Immunol. 135*, 3712-3718.

Hayday, A., Diamond, D., Tanigawa, G., Heilig, J., Folsom, V., Saito, H., and Tonegawa, S. (1985). Unusual features of the organization and diversity of T-cell receptor α chain genes. *Nature 316*, 828-832.

Heber-Katz, E., Schwartz, R. H., Matis, L. A., Hannum, C., Fairwell, E., Appella, E., and Hansburg, D. (1982). Contribution of antigen-presenting cell major histocompatibility complex gene products to the specificity of antigen induced T-cell activation. *J. Exp. Med. 155*, 1086-1099.

Katz, D. H., Hamaoka, T., and Benacerraf, B. (1973). Cell interactions between histoincompatible T and B lymphocytes. II. Failure of physiologic cooperative interactions between T and B lymphocytes from allogeneic donor strains in humoral response to hapten-protein conjugates. *J. Exp. Med. 137*, 1405-1418.

Kim, S., Davis, M., Sinn, E., Patten, P., and Hood, L. (1981). Antibody diversity: Somatic hypermutation of rearranged V_H genes. *Cell 27*, 573-581.

Kindred, B., and Shreffler, D. C. (1972). H-2 dependence of cooperation between T and B cells in vivo. *J. Immunol. 109*, 940-943.

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

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

Malissen, M., McCoy, C., Blanc, D., Trucy, J., Devaux, C., Schmidtt-Verhust, A.-M., Fitch, F., Hood, L., and Malissen, B. (1986). Direct evidence for chromosomal inversion during T-cell receptor β gene rearrangements. *Nature 319*, 28-33.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). Molecular Cloning, A Laboratory Manual. (Cold Spring Laboratory, Cold Spring Harbor, New York).

Matis, L. A., Hedrick, S. M., Hannum, C., Ultee, M. E., Lebwohl, D., Margoliash, E., Solinger, A. M., Lerner, E. A., and Schwartz, R. H. (1982). The T-lymphocyte response to cytochrome c. III. Relationship of the fine specificity of antigen recognition to major histocompatibility complex genotype. *J. Immunol. 128*, 2439-2446.

McKean, D., Huppi, K., Bell, M., Staudt, L., Gerhard, W., and Weigert, M. (1984). Generation of antibody diversity in the immune response of BALB/c mice to influenza virus haemaglutinin. *Proc. Natl. Acad. Sci. USA 81*, 3180-3184.

Mengle-Gaw, L., Conner, S., McDevitt, H. O., and Fathman, C. G. (1984). Gene conversion between murine class II major histocompatibilty loci. *J. Exp. Med. 160*, 1184-1194.

Mengle-Gaw, L., and McDevitt, H. O. (1985). Predicted protein sequence of the murine I-E^{s}_{β} polypeptide chain from cDNA and genomic clones. *Proc. Natl. Acad. Sci. USA 82*, 2910-2914.

Patten, P., Yokota, T., Rothbard, J., Chien, Y.-H., Arai, K.-I., and Davis, M. M. (1984). Structure, expression and divergence of T-cell receptor β chain variable regions. *Nature 312*, 40-46.

Pech, M., Hochlt, J., Schnell, H., and Zachau, H. G. (1981). Differences between germline and rearranged immunoglobulin Vκ coding sequences suggest a localized mutation mechanism. *Nature 291*, 668-670.

Perlmutter, R. M., Crew, S. T., Douglas, R., Sorensen, G., Johnson, N., Nivera, N., Gearhart, P., and Hood, L. (1984). The generation of diversity in phosphorylcholine-binding antibodies. *Adv. Immunol. 35*, 1-37.

Rosenthal, A. S., and Shevach, E. M. (1973). Function of macrophages in antigen recognition by guinea pig T lymphocytes. *J. Exp. Med. 138*, 1194-1212.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain termination inhibitors. *Proc. Natl. Acad. Sci. USA 74*, 5463-5467.

Schilling, J. (1981). Antibody Diversity. Ph.D. Thesis, California Institute of Technology.

Schilling, J., Clevinger, B., Davie, J. M., and Hood, L. (1980). Amino acid sequence of homogeneous antibodies to dextran and DNA rearrangements in heavy chain V region gene segmens. *Nature 283*, 35-40.

Schwartz, R. H. (1985). T-lymphocyte recognition of antigen in association with gene products of the major histocompatibility complex. *Ann. Rev. Immunol. 3*, 237-261.

Siegelman, M., and Capra, J. D. (1981). Complete amino acid sequence of light chain variable regions derived from five monoclonal anti-p-azophenylarsonate antibodies differing with respect to a cross-reactive idiotype. *Proc. Natl. Acad. Sci. USA 78*. 7679-7683.

Solinger, A. M., Ultee, M. E., Margoliash, E., and Schwartz, R. H. (1979). T-lymphocyte response to cytochrome c. I. Demonstration of a T-cell heteroclitic proliferative response and identification of a topographic antigenic determinant on pigeon cytochrome c whose immune recognition requires two complementing major histocompatibility complex-linked immune response genes. *J. Exp. Med. 150*, 830-848.

Strauss, E. C., Kobori, J. A., Siu, G., and Hood, L. (1986). Specific primer directed DNA sequencing. *Anal. Biochem.*, in press.

Tonegawa, S. (1983). Somatic generation of antibody diversity. *Nature 302*, 575-581.

Winoto, A., Mjolsness, S., and Hood, L. (1985). Genomic organization of the genes encoding mouse T-cell receptor α chain. *Nature 316*, 832-836.

Yoshikai, Y., Clark, S. P., Taylor, S., Sohn, U., Wilson, B., Minden, M., and Mak, T. W. (1985). Organization and sequences of the variable, joining and constant region genes of the human T-cell receptor α chain. *Nature 316*, 837-840.

Zinkernagel, R. M., and Doherty, P. C. (1974). Restriction of in vitro T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semi-allogeneic system. *Nature 248*, 701-702.

Figure 1. Sequences of Peptide Antigens Used in this Study

The sequences of the peptides used in this study (fly, DASp, Sp, DAK^{100}, DAE^{100}, DAK^{99}) are compared to the sequence of pigeon cytochrome c C-terminal residue 81-104. Note that an N-terminal cysteine has been added to the fly sequence. — denotes identity and Δ denotes a deletion. Deletion was introduced for convenience during peptide synthesis.

81
91
100
104
Pigeon Cyto.c I F A G I K K K A E R A D L I A Y L K Q A T A K
fly C L P N G S Δ
DASp A N E Δ Δ Δ Δ
Sp A N E Δ Δ Δ Δ
DAK100 A N E Δ Δ Δ Δ K Δ
DAE100 A N E Δ Δ Δ Δ E Δ
DAQ99 A N E Δ Δ Δ Δ Q Δ

Figure 2. Northern Blot Analysis Using the $V_{\alpha 11.1}$ Gene Segment as a Probe

Northern blot analysis of total RNAs from five different I-E^k and six different I-E^s-restricted/DASp-specific T-helper hybridomas are shown. The probe was a 2.2 kb Hind III fragment of the $V_{\alpha 11}$ rearranged gene from hybridoma AN6.2.

I-E^K
I-E^S
AN4.4
AN6.2
AN14.4
BC15.1
BC37.5
AP11.2
AP15.2
AP7.10
AP10.10
AR8.1
AR10.1
9x1.7kb

Figure 3. Southern Genomic Blot Using 3.6 kb Kpn I/Bam HI Fragment from the $V_{\alpha}11.1$ Gene Segment Flanking Region as a Probe

Southern blot analysis of six cytochrome-c specific T-helper hybridomas. The genomic blot of hybridoma AP15.2 is in the germline configuration, hybridomas AN6.2 and BC15.1 show 15 kb rearranged bands that were also detected by the Jα84 probe (data not shown) and hybridomas AN14.4 and BC37.5 show 6 kb rearranged bands that were subsequently cloned. Hybridoma AR8.1 shows a 6.6 kb rearranged band.

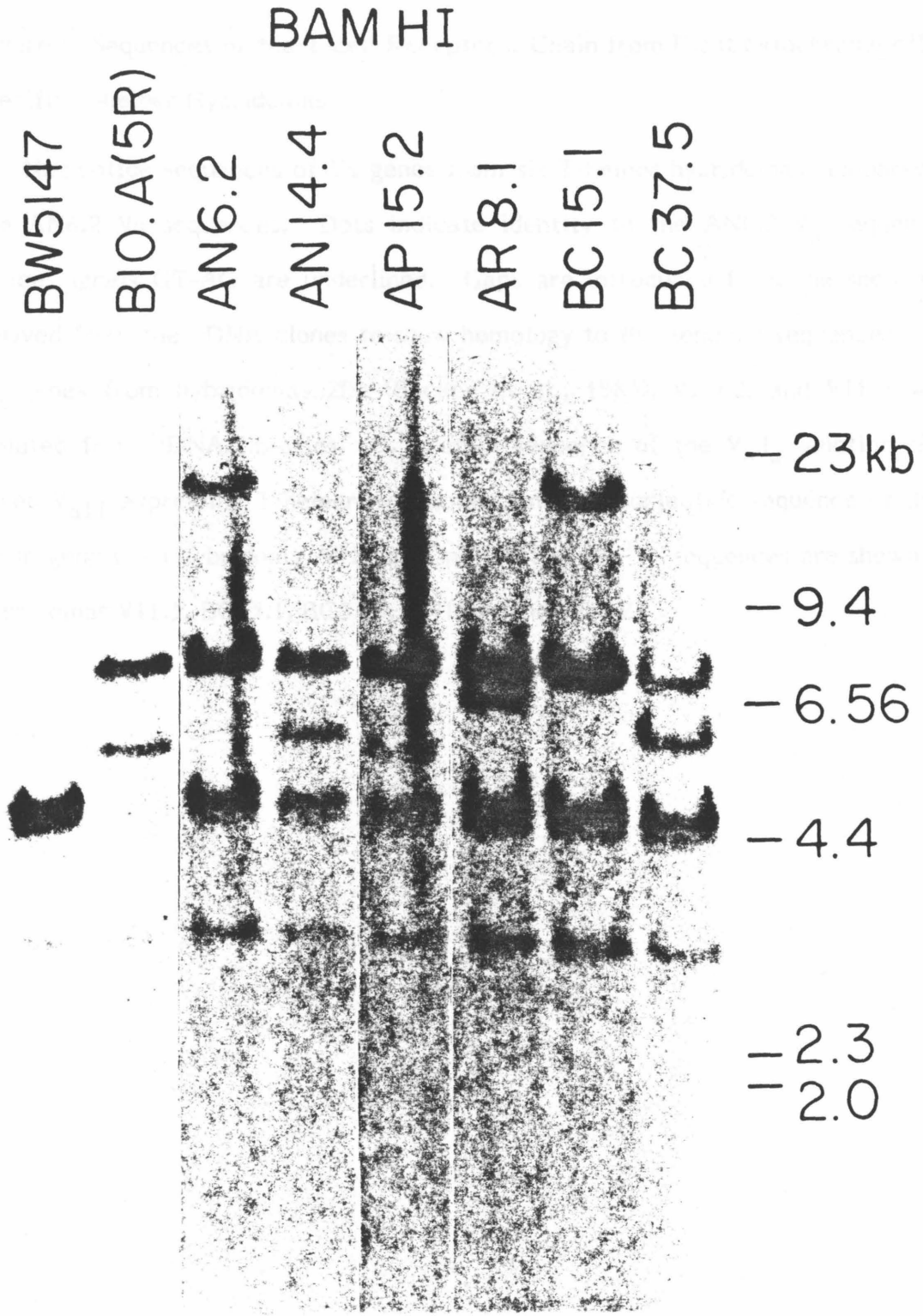

BAM HI
BW5147
BIO.A(5R)
AN6.2
AN14.4
AP15.2
AR8.1
BC15.1
BC37.5
23kb
9.4
6.56
4.4
2.3
2.0

Figure 4. Sequences of the T-cell Receptor α Chain from Eight cytochrome c/I-E-specific T-helper Hybridomas

A. Nucleotide sequences of Vα genes from six T-helper hybridomas compared to the AN6.2 Vα sequences. Dots indicate identity to the AN6.2 V_α sequences. Splice signals GT-AG are underlined. Gaps are introduced from the sequences derived from the cDNA clones to show homology to the genomic sequences. The V_α genes from hybridomas 2B4 (Becker et al., 1985), V1.9.2, and V11.5 were isolated from cDNA libraries. B. Protein sequence of the $V_\alpha J_\alpha$ junction from seven $V_{\alpha 11}$ expressing T-helper hybridomas and the nucleotide sequence of the α chain gene from hybridoma AP11.2. Only part of the V_α sequences are shown for hybridomas V11.5, BC15.1, BC37.5, AN14.4, and V1.9.2.

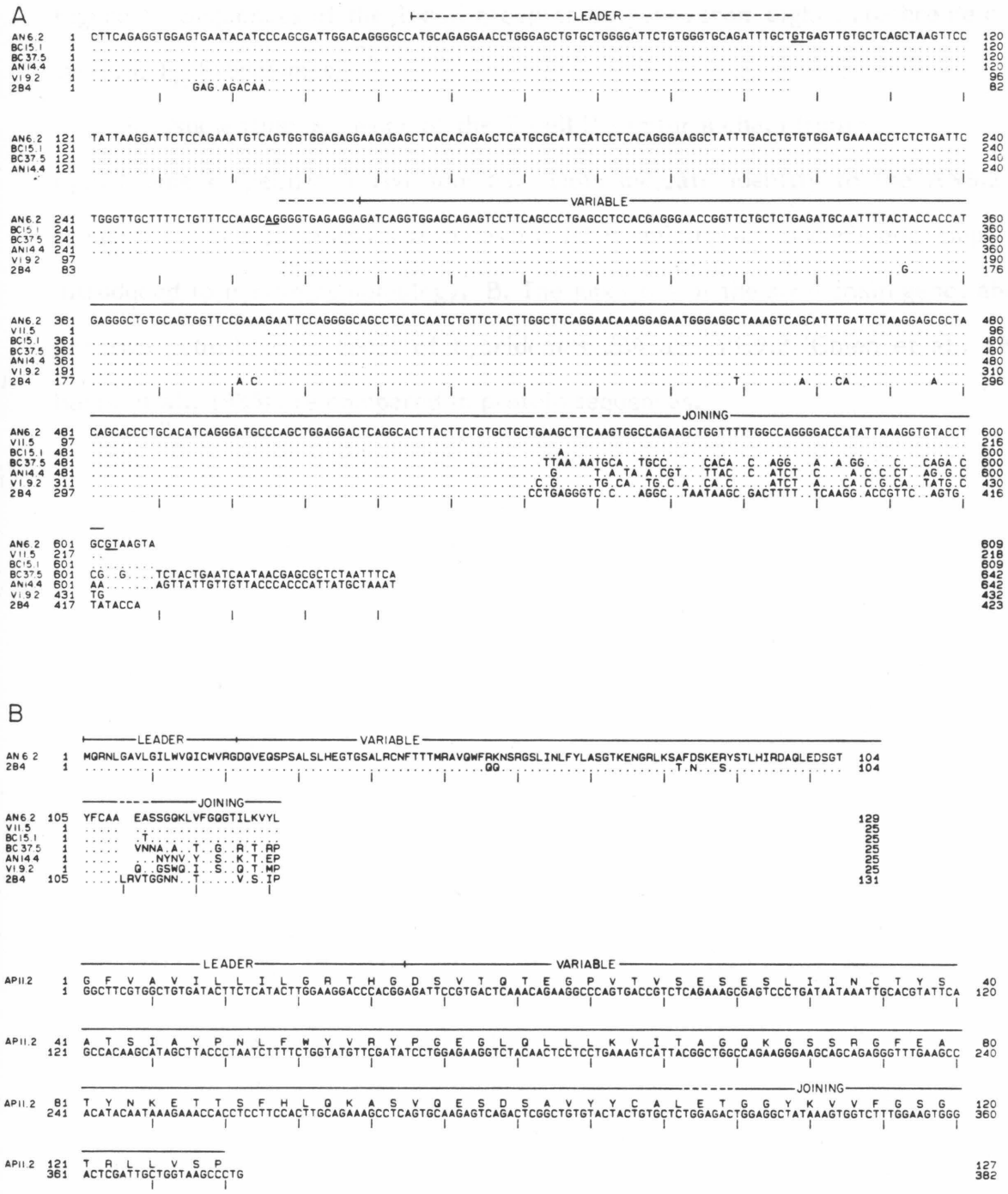
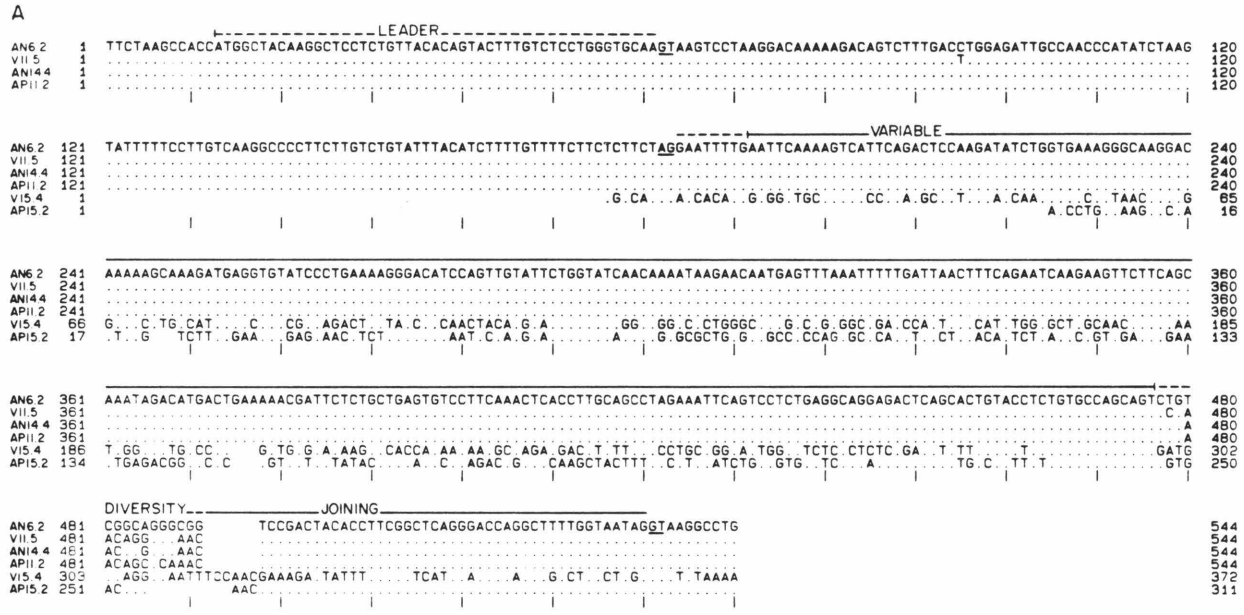

A
LEADER
VARIABLE
JOINING
AN6.2
BC15.1
BC37.5
AN14.4
VI9.2
284
VII.5
B
LEADER
VARIABLE
JOINING
AN6.2
284
VII.5
BC15.1
BC37.5
AN14.4
VI9.2
APII.2

Figure 5. Sequences of the T-cell receptor V_β genes from eight cytochrome c/I-E specific T_H hybridomas.

A. Nucleotide Sequence of the T-cell Receptor β chain from Six cytochrome c-specific T Hybridomas. Dots indicate identity to the AN6.2 V_β sequences. The canonical sequences for RNA splicing are underlined. Gaps are introduced to maximize homology. B. The junctions of the six β chain genes above and two published V_β genes of hybridomas 2B4 and V1.9.2 (Chien et al., 1984; Barth et al., 1985) are compared in protein sequences.


A
LEADER
VARIABLE
DIVERSITY
JOINING
AN6.2
VII.5
ANI4.4
API1.2
VI5.4
API5.2

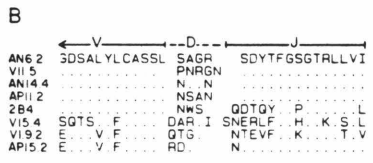

B
V
D
J
AN6.2
VII.5
ANI4.4
API1.2
2B4
VI5.4
VI.9.2
API5.2

Figure 6. The Fine Specificity of Pigeon Cytochrome c-specific T Hybridomas Using B10.A Antigen Presenting Cells

Hybridomas cells were cultured with varying concentrations of antigens (the whole molecule of pigeon cytochrome c was used in this study) and B10.A spleen cells as source of antigen presenting cells. After 40 hr of culture, supernatants were transferred to the IL2-dependent cell line HT-2 to measure IL-2 activity. After a further 16 hr of culture, ^{3}H-thymidine was added to the secondary culture. Hybridomas: A, AN4.4; B, AN14.4; C, BC37.5; D, V11.5; E, AN6.2. Antigens: DASp; pigeon cytochrome c; fly; Sp; DAK^{100}; DAE^{100}. The Y coordinates are different for different hybridomas to emphasize the relative responses for different peptides.

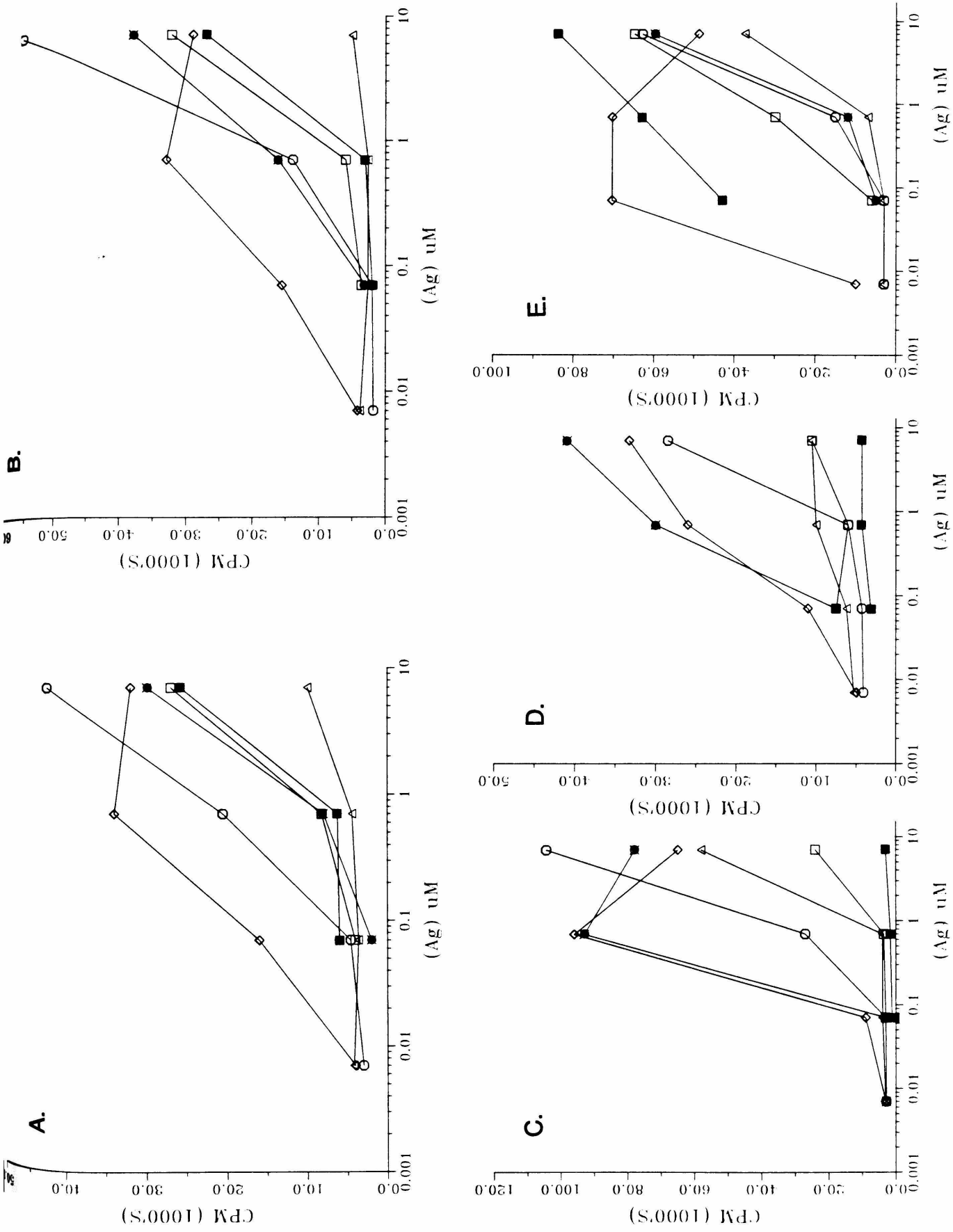

A.
B.
C.
D.
E.
CPM (1000'S)
(Ag) uM
0.0
10.0
20.0
30.0
40.0
50.0
60.0
80.0
100.0
120.0
0.001
0.01
0.1
1
10

Figure 7. Comparison of Fine Specificity of T Hybridomas

A, AP11.2 on B10.S(9R) APC; B, V15.4 on B10.A APC; C, BC15.1 on B10.A APC. In separate experiments AP11.2 was found to be non-reactive with the DAK^{100} peptide, and the fly cytochrome c gave comparable response as the Sp peptide. See Figure 6 for symbols representing the antigens.

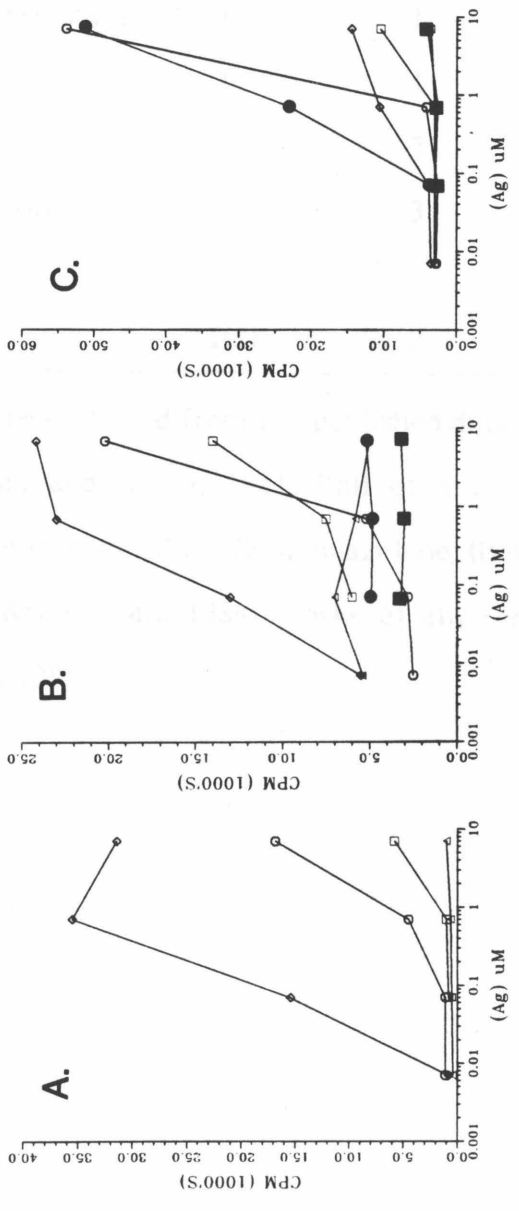

A.
B.
C.
CPM (1000'S)
(Ag) uM
0.0
5.0
10.0
15.0
20.0
25.0
30.0
35.0
40.0
50.0
60.0
0.001
0.01
0.1
1
10

Table 1. Numbers of Different Immunoglobulin V_H and V_L Genes Used in the Secondary Response to Several Antigens

Antigens	V_H	J_H	V_L	J_L
Arsenate Id^{CR} determinant	1	1	1	1
Phosphorylcholine	1	1	3	1
Oxazolone	5	3	4	4
Influenza haemaglutinin Sb determinant	3	3	1	3
α1,3 Dextran	1(?)	2	1	1

The data above were compiled from the published data on the antibody responses to arsenate (Siegelman and Capra, 1981; Ball et al., 1983), phosphorylcholine (for review, see Perlmutter et al., 1984), oxazolone (Berek et al., 1985), influenza haemaglutinin (McKean et al., 1984; Clarke et al., 1985) and dextran (Schilling et al., 1980; Schilling, 1981).

Table 2. T_H Hybridomas, Their Strains of Origin, Their Restricting Elements and the Antigens Used for Immunizations

T_H hybridomas	Strain	Restricting elements	Immunogens	
			1 (in vivo)	2 (in vitro)
V1.9.2*, V11.5, V15.4	B10.A(5R)	$E_\alpha^k E_\beta^b$	DASp	DASp
AN4.4, AN6.2, AN14.4	B10.A	$E_\alpha^k E_\beta^k$	Sp	DASp
AP7.10, AP10.10 AP11.2, AP15.2	B10.S(9R)	$E_\alpha^k E_\beta^s$	DASp	DASp
AR8.1, AR10.1	B10.SM	$E_\alpha^v E_\beta^v$	DASp	DASp
BC15.1, BC37.5	B10.A	$E_\alpha^k E_\beta^k$	Pigeon cyt. c	Sp

*Although hybridoma V1.9.2 was originally obtained from a mouse immunized with the DASp peptide, it was found after fusion and cloning to react with the AmDASp peptide (acetimidyl form of the DASp peptide) but not with the DASp peptide.

Table 3. Hybridomas, Restricting Elements and Gene Segments

Hybridomas	Restriction	V_α*	J_α	V_β	D_β	J_β
AN6.2	I-E^k	11.1	84	3	1	1.2
AN14.4	"	11.1	14.4	3	1 or 2	1.2
BC15.1	"	11.1	84	N.D.	N.D.	N.D.
BC37.5	"	11.1	37	14	N.D.	N.D.
2B4	"	11.2	2B4	3	2	2.5
V11.5	"	11.1	84	3	1	1.2
V1.9.2	"	11.1	28	1	1	1.1
V15.4	"	11	N.D.	8.3	1	1.4
AP11.2	I-E^s	4.3	11.2	3	1	1.2
AP15.2	"	N.D.	N.D.	1	1 or 2	1.2

N.D.: Not done (cloning and sequencing of the corresponding genes have not been carried out).

*The designation of the V_α, J_α, and V_β gene segment nomenclature is essentially as described before (Arden et al., 1985; Barth et al., 1985; Kronenberg et al., 1986). The two distinct members of the $V_{\alpha 11}$ subfamilies are denoted $V_{\alpha 11.1}$ and $V_{\alpha 11.2}$. The V_α gene of hybridomas AP11.2 is denoted $V_{\alpha 4.3}$ to distinguish it from the two published V_α sequences belonging to the $V_{\alpha 4}$ subfamilies (Arden et al., 1985). The J_α gene segments used by the AN14.4 and AP11.2 hybridomas have not been reported before and are denoted here as $J_{\alpha 14.4}$ and $J_{\alpha 11.2}$ gene segments, respectively. The information for the 2B4 V genes and the V1.9.2 V_β gene has been previously published (Chien et al., 1984; Barth et al., 1985; Becker et al., 1985).

Table 4. V Gene Segment Usage in T_H Hybridomas Specific for the C-terminal Peptide of Cytochrome c and E Molecule

MHC restriction of the T_H hybridomas:	V_α gene segment usage*		
	$V_{\alpha 4}$	$V_{\alpha 11}$	Others
I-E^k	0	8	1
I-E^s	1	3	2

	V_β gene segment usage$^\Delta$						
	$V_{\beta 1}$	$V_{\beta 2}$	$V_{\beta 3}$	$V_{\beta 8.1/8.2/8.3}$	$V_{\beta 8.3}$	$V_{\beta 14}$	Others (x)$^+$
I-E^k	1	1	4	1	1	2	2
I-E^s	1	0	1	2		0	2

*Data were from the sequence analyses of the $V_{\alpha 4}$ gene from one E^s restricted hybridoma, and seven $V_{\alpha 11}$ genes from the E^k restricted hybridomas. The remaining data were derived from Northern analyses of RNAs from seven additional T_H hybridomas.

$^\Delta$Data were from the sequence analyses of two $V_{\beta 1}$ genes and five $V_{\beta 3}$ genes and one $V_{\beta 8.3}$ gene as well as Northern blot analyses from the RNAs of ten additional T_H hybridomas. The $V_{\beta 2}$, $V_{\beta 3}$, $V_{\beta 4}$, $V_{\beta 5}$, $V_{\beta 8}$ and $V_{\beta 14}$ gene segments were used as probes.

$^+$x denotes the usage of different V_β gene segments other than the $V_{\beta 2}$, $V_{\beta 3}$, $V_{\beta 4}$, V_5, $V_{\beta 8}$ and $V_{\beta 14}$ gene segments.

Chapter Six

THE FUTURE

Molecular Immunology has been an exciting field for the past few years. The availability of monoclonal antibodies, the ability of growing T-cell lines in culture and to transfect T or B cells *in vitro* with ease and the cloning of a lot of biologically important molecules have advanced the field of immunology to a new dimension that no one has ever dreamed of. We are now in a position to be able to redefine and study the immune system step by step in great detail with confidence. Cloning of the immunoglobulin, class I and class II MHC molecules and the T-cell receptor molecules in particular have provided us with an incredible amount of invaluable information on how the immune system works and have cleared up a lot confusion that existed in immunology. Molecular Immunology is moving into two general directions: the study of the regulation of the immune system and the structure function study of the T-cell receptor molecules, in particular, on how the T-cell receptor molecules recognize antigen and MHC molecule together and how the thymus education works.

REGULATION

There are a lot of interesting regulation problems in the immune system, only some of which are mentioned below:

Questions:

How does the allelic exclusion work in B-cell and T-cell receptor? How does the recombinase enzyme work? Are there any other cells/tissues that express the recombinase? This last question relates to the question of whether any other biological system employs the strategy that has been used so successfully in the immune system to generate enormous diversity. How does the switching enzyme work? How does the somatic hypermutation in B cells occur and why would not it happen in the T-cell receptor V gene?

Approach:

The only approach that will undoubtedly answer lots of the questions above,

will be the development of *in vitro* systems for the VDJ or VJ recombination, immunoglobulin switching and the immunoglobulin hypermutation. Any other approach such as cloning by subtractive probe or DNA transfection will run into the problems of the inability to prove whether the clone encodes the appropiate molecules and whether there is actually more than one enzyme that act together to produce the observed effects. The difficulty in developing the *in vitro* system, however, lies in finding a simple and reliable assay system to detect each of the above reactions once they occur *in vitro*. Another difficulty is the establishment of cell lines that can switch or hypermutate immunoglobulin genes all the time. Cell lines that rearrange κ gene continuously have been developed. Once the *in vitro* system has been developed, enzymes responsible for the recombination, switching and hypermutation can be purified and studied in detail. Amino acid sequences of the corresponding molecules can be obtained and the amino acids information can be used to clone the corresponding gene. Finally, the detailed mechanism of VDJ or VJ recombination, immunoglobulin switching and somatic hypermutation can be studied *in vitro* using the purified enzymes.

Questions:

Is there any enhancer in the T-cell receptor α chain? If the α chain enhancer exists, how many are there? If there is only one enhancer, is there any special feature to this enhancer allowing it to act more than 60 kb away? Finally, how does an enhancer work and what enzyme(s) is (are) involved in the process?

Approach:

The question on the T-cell receptor α chain enhancer(s) can be approached easily by transfecting a the rearranged V_α gene into a T-cell. The effect on the level of transcription of the V_α gene by various restriction fragments isolated from the 60 kb of DNA 5' to the C_α gene can be measured. Once an enhancer is identified, sequencing work can be performed and an assay comparing the α chain

enhancer to the immunoglobulin enhancer can be done.

The question of how the enhancer works is a more difficult problem to approach. Again, I think the ultimate solution is the ability to see the enhancer effect *in vitro* so that the enzymes involved in the process can be purified biochemically and the effect of enhancer on various *in vitro* manipulated DNA can be studied.

Structure function study of T-cell receptor molecule

Questions:

Given a V_α gene segment, VJ α chain junctional sequences, a J_α gene segment, a V_β gene segment, VDJ β chain junctional sequences, and a J_β gene segment to constitute functional T-cell receptors V gene(s), what is(are) the important part(s) in determining the MHC restriction, antigen specificity and antigen fine specificity?

Approach:

The cytochrome c T-helper system that I have analyzed offers a golden opportunity to study the above questions. I have suggested that the V_α gene segment may play a crucial role in determining the antigen specificity; this can be tested by performing gene transfer of various combination of V_α and V_β genes into another functional T cell and by studying the effect of the different combinations on the MHC restriction and antigen fine specificity of the transfected cell.

Examples:

1. If the $V_{\alpha}11.1$ gene in combination with other V_β gene ($V_{\beta 1,3,8,14}$) is responsible for the reactivity of the T cell hybridoma to the C terminal peptide of cytochrome c containing lysine 99, then the V_α11.1 gene isolated from the V1.9.2 hybridoma is not the right V_α gene for the AmDASp reactivity of the V1.9.2 hybridoma. Since the V1.9.2 hybridoma originates from a DASp immunized mouse, it is possible that the V_α11.1 gene from hybridoma V1.9.2 in combination with the

V_β gene from V1.9.2 will give the DASp reactivity in a transfected cell. The AmDASp reactivity of hybridoma V1.9.2 is in turn due to the contribution of the T-cell receptor molecules of the BW5147 fusion partner. This can also be tested using gene transfer of a BW5147 α chain gene in combination with the V1.9.2 V_β gene, or a BW5147 β chain gene in combination with the V1.9.2 V_α chain gene. If my hypothesis is correct, then the T cell transfected with the BW5147 V_α gene and the V1.9.2 V_β gene will endow the T cell with a new reactivity to the AmDASp peptide. This will also open an exciting possibility to study the interaction of V protein *in vitro*. Using purified V1.9.2 and BW5147 T-cell receptor molecules, one can ask the questions of what determines preferential association of V proteins, as it might be the case here, and how the V proteins interact with one another.

2. Hybridoma AP11.2 employs a $V_{\alpha 4.3}$ gene segment, a $J_{\alpha 11.2}$ gene segment and a $V_{\beta 3}$ gene segment with a $J_{\beta 1.2}$ gene segment. This hybridoma is interesting because it uses similar β chain gene as the hybridoma AN6.2 except for the junctional sequence. Hybridoma AP11.2 is restricted to the E^s MHC molecule, whereas hybridoma AN6.2 is restricted to the E^k MHC molecule. Obviously, either the α chain or the junctional sequence of β chain of the T-cell receptor is crucial for the differences of MHC restriction between the two hybridomas. To address the issue of which element(s) is (are) important in determining MHC restriction of the T-cell hybridoma, the "exon" shuffling experiment should be performed to construct V_α genes consisting of a $V_{\alpha 11.1}$ gene segment, a VJ α chain junctional sequence of hybridoma AP11.2 and $J_{\alpha 11.2}$ gene segment and the reciprocal product; the $V_{\alpha 11.1}$ gene segment, VJ α chain junctional sequences of hybridoma AN6.2 and $J_{\alpha 11.2}$ and the reciprocal product. The constructs should be transfected in all kinds of combination with the β chain gene from hybridoma AP11.2 and AN6.2. The transfected cells will then be tested for reactivity to the C-terminal peptide of cytochrome c in the context of either the E^s or E^k molecules. This approach will

hopefully pinpoint one or two elements that play crucial role in determining MHC restriction.

Questions:

What is the three-dimensional structure of the T-cell receptor molecules and how does this relate to the recognition of the antigen/MHC molecules by the T-cell receptor molecule and the education for recognition of self MHC molecule during thymus development?

Approach:

Purify the soluble T-cell receptor molecules in huge quantities, crystallize them and solve the three-dimensional structure of the T-cell receptor in solution. The expression of T-cell receptor protein can be done either in the insect system or the chinese hamster ovary system to ensure proper glycosylation of the molecules. Once the purified molecule is available, binding studies to the MHC molecule alone or the antigen/MHC molecule can be done. One possible difficulty will be the lack of binding due to the fact that all three elements are not presented on a cell membrane. If this is the case, a liposome system of purified membrane T-cell receptor and monolayer system of antigen/ MHC molecule should be used. This will, we hope, solve a lot of the longstanding questions of T-cell receptor recognition and provide a good working model of how thymus education works.

In summary, I think a lot of the molecular immunological questions in the near future can be approached by the development of the *in vitro* system and by the *in vitro* physical chemistry studies of the immunologically important molecules.

APPENDIX A

ISOLATION OF THE T-CELL RECEPTOR α CHAIN cDNA

Project

To isolate the cDNA encoding the T-cell receptor α chain

Strategy

1. Isolate T-cell specific cDNA clones by the subtractive method.
2. Using the inserts isolated from the T-cell specific clones, do genomic blots on the T-cell DNA to detect DNA rearrangement.
3. Any cDNA clone that is T-cell specific and shows DNA rearrangment is a candidate for the T-cell receptor α chain cDNA and will be further characterized.

Results

I have constructed a cDNA library from a T-cell hybridoma specific for the C-terminal peptide of cytochrome c (V1.9.2) in a λgt 10 vector. The library contains around 200,000 clones with average cDNA inserts around 1.3 kb. To amplify the signals to be detected and to avoid rescreening of the positive clones, I have picked 10,000 cDNA clones and grew them as an array in petri dishes containing agarose. The clones were then screened with one million cpm/ml of subtractive probes. The probes consisted of a subtractive probe of T-cell cDNA minus B-cell RNA and B-cell cDNA minus macrophage RNA. The B-cell minus the macrophage probe was included as a control to eliminate any positive clones due to the presence of repeat elements in the cDNA clones. 64 clones were isolated this way. Genomic blot analyses were carried out using the inserts from these 64 clones. The germline mouse DNA, the BW5147 fusion partner DNA and the V1.9.2 T-cell hybridoma DNA digested by EcoRI, BamHI, Hind III and PvuII enzymes were included in the genomic blots. DNA rearrangments were detected for only one clone (α 60); the clone, however, is a T-cell receptor β chain cDNA by hybridization to the probe isolated

by Tak Mak' group. When Mark Davis' and Tonegawa's groups published their T-cell receptor α chain cDNA sequences, I used oligonucleotide specific for the α chain sequence as a probe to hybridize to the 64 T-cell specific cDNA clones. One of them hybridized very strongly to the probe, indicating that I did have the T-cell receptor α chain clones among the pool of T-cell specific clones. The clone, however, contains only the constant region of the T-cell receptor α chain. This is the reason I failed to detect any DNA rearrangment on the genomic blot analysis using the corresponding clone (see Chapters One and Four for explanation). Below is a summary of the interesting T-cell specific clones that I have obtained during this project:

Clones α1, α4=α6=α61, α19=α21=α26=α28=α34=α49=α63 and α55 are four different multigene families as they showed anywhere from 10 to 50 bands in genomic blot analyses.

Clone α23 encodes the T-cell receptor α chain.

Clone α60 encodes the T-cell receptor β chain.

Clone α11 is very interesting because it is extremely polymorphic.

APPENDIX B

MOLECULAR CHARACTERIZATION OF THE RECOMBINATION REGION OF SIX MURINE MAJOR HISTOCOMPATIBILITY COMPLEX (MHC) I-REGION RECOMBINANTS

Published in

Journal of Molecular and Cellular Immunology **1**, 125-131, 1984

Molecular Characterization of the Recombination Region of Six Murine Major Histocompatibility Complex (MHC) I-Region Recombinants

Joan A. Kobori, Astar Winoto, Janet McNicholas[1]*, and Leroy Hood
Division of Biology, California Institute of Technology, Pasadena, California 91125; and [1]Department of Biological Science, Stanford University, Stanford, California 94305, USA

Précis

Using Southern DNA hybridization techniques, restriction enzyme site polymorphisms have been used to correlate the molecular maps of the murine major histocompatibility complex (MHC) I region with the genetic map derived from analyses of recombinant mouse strains. The data indicated that the DNA that maps between the I-A and I-E subregions is limited to 3.4 kilobases (kb) and includes the 3′ end of the E_β gene. According to classical genetic mapping by recombinational analysis of serological markers, this region should encode the I-B and I-J subregions. These observations are surprising in two respects. First, 3.4 kb is a small amount of DNA to encode even one complete murine gene. Second, this region, which putatively encodes the I-J gene, appears to reside at least partially within the E_β gene.

To analyze these apparent paradoxes, further, we cloned the 3.4-kb region in question from six I-region combinant strains [B10.A(3R), B10.a(5R), B10.A(4R), B10.GD, B10.HTT, and B10.S(9R)] and four strains used in the derivation of the recombinants (B10.D2, B10.A, C57BL/10, and ASW) into a lambda phage vector. By direct restriction enzyme mapping of polymorphic sites, we have confirmed the previously identified boundaries of the I-A and I-E subregions and have narrowed the estimate of the distance between these subregions to approximately 2.0 kb of DNA. This 2.0-kb region encompasses part of the intron between the first- (β1) and second-domain (β2) exons and the second-domain exon (β2) of the E_β gene. Thus, we can conclude that the I-B and I-J subregions as mapped by immunogeneticists are contained within 2-kb of DNA and must be present entirely within the E_β gene. A serological difference between the B10.A(3R) and B10.A(5R) strains has been observed that has been mapped to the I-J subregion. However, at the molecular level the DNA between the I-A and I-E subregions of these strains seems to be identical by restriction map analysis. Our data also suggest that the crossover points in all of the I-region recombinants examined may have occurred within approximately 2 kb of DNA. These observations have two implications. (1) The I-B and I-J subregions may not be encoded between the I-A and I-E subregions. (2) A hot spot for recombination has been observed in the I region; hence, the genetic map distances in the MHC that were assigned on the assumption that recombination is random must be viewed with caution.

Present address: Department of Microbiology and Immunology, University of Illinois, Box 6998, Chicago, Illinois 60680
Address correspondence to: Joan A. Kobori

Introduction

The major histocompatibility complex (MHC) of the mouse encodes three classes of proteins involved in immune responses [1, 2]. The class I molecules include the transplantation antigens, cell-surface glycoproteins that act as restriction elements in the response of cytotoxic T cells to virally infected cells. The class II molecules or I-region-associated (Ia) antigens are cell-surface glycoproteins that are present on bone-marrow-derived cells (B cells, T cells, and macrophages) and play a role in cellular interactions regulating immune responses. The class III molecules include several serum components of the complement pathway.

Five I subregions have been defined by serological recombinational analysis: I-A, I-B, I-J, I-E, and I-C (Fig. 1) [3]. The boundaries of the subregions are defined operationally as the recombination crossover positions in H-2 recombinant strains (Table 1). The molecular cloning of the I region of the MHC from BALB/c mice provided DNA hybridization probes that were used to compare the

J Mol Cell Immunol (1984) 1:125–131

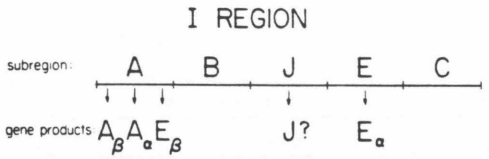


Fig. 1. Schematic representation of the I region of the major histocompatibility complex of the mouse

Table 1. Recombinant H-2 haplotypes defining the subregions of the I region[a]

	Haplotype origin of region or subregion						
		I region					
Strain	K	A	B	J	E	C	D
A.TL	*s*	*k*	*k*	*k*	*k*	*k*	*d*
B10.AQR	*q*	*k*	*k*	*k*	*k*	*d*	*d*
B10.A(4R)	*k*	*k*	*b*	*b*	*b*	*b*	*b*
B10.GD, D2.GD	*d*	*d*	*b*	*b*	*b*	*b*	*b*
B10.A(5R)	*b*	*b*	*b*	*k*	*k*	*d*	*d*
B10.A(3R)	*b*	*b*	*b*	*b*	*k*	*d*	*d*
B10.HTT	*s*	*s*	*s*	*s*	*k*	*k*	*d*
B10.A, A	*k*	*k*	*k*	*k*	*k*	*d*	*d*
C3H.OL	*d*	*d*	*d*	*d*	*d*	*d*	*k*

[a] Adapted from reference 3

organization of the I region in recombinant congenic strains. Using the Southern DNA blotting technique with restriction-enzyme-digested genomic DNA, Steinmetz et al. constructed a partial restriction map for each I region recombinant [4]. Their data indicated that the I-A and I-E subregions were separated by at most 3.4 kilobases (kb) of DNA and that, accordingly, this chromosomal region should encompass the I-B and I-J subregions. The I-B subregion was defined by its effect on regulating immune responsiveness toward several antigens in certain inbred strains of mice [5, 6]. However, the requirement for the I-B subregion has been refuted by Baxevanis et al. on the basis of data demonstrating interaction between the I-A and I-E subregions [7]. The I-J subregion appears to encode a polypeptide, the I-J chain, that has been identified both with complex antisera and, more recently, with monoclonal antibodies [8–11]. The I-J gene product is detected on suppressor T cells by means of serological reagents and is associated with soluble inhibitory factors released by suppressor T cells [8–16]. The I-J gene has been of considerable interest because it appears to encode a portion of the suppressor T cell receptor for antigen. The fact that the I-J gene apparently was confined to such a small chromosomal region in and around the E_β gene led us, in the studies reported below, to isolate and clone this region from the parental and I-region-recombinant strains.

Results and Discussion

Isolation of the E_β *Gene Region from Parental and I-Region-Recombinant Strains*

Using DNA hybridization probe 4, which is specific for the 3′ end of the E_β^d gene, Steinmetz et al. [4] showed by genomic Southern blotting that the restriction endonuclease EcoRI revealed polymorphism in the DNA from various mouse strains examined. The characteristic fragments observed in the parental strains were 12.2 kb for the *d*, *k*, and *s* haplotypes and 6.2 kb for the *b* haplotype.[1] Using this polymorphism, we digested the DNAs of four parental haplotype and six I-region-recombinant mouse strains with EcoRI, isolated the restriction fragments of E_β size from an agarose gel, and cloned them into the phage vector λgt7. This procedure provided an approximately 20-fold enrichment for the desired clones. The hybridization probe used to screen these partial libraries was again probe 4 of Steinmetz et al. [4]. DNA was isolated from the positive hybrid phage for subsequent analysis.

Determination of Haplotype-Specific Restriction Enzyme Site Polymorphism in the Hybrid Phage

The DNA of hybrid phage containing the desired region of the E_β genes of parental haplotypes were digested with as many six-base recognition site restriction endonucleases as could be obtained and readily used to map the DNA. Restriction enzyme site polymorphisms distinguishing the various haplotypes were determined and a partial restriction map for the E_β region was constructed (Fig. 2). Detailed restriction analyses of the I region from cosmid clones of the *d* and *k* haplotypes suggest colinearity of the genome in the E_β gene region [M. Steinmetz and L. Hood, unpublished observations]. Thus, we assume that the *s* and *b* haplotypes also will be colinear in this region. This assumption is supported by the facile alignment of the restriction maps from the four parental haplotypes with each other. Noteworthy is the fact that except for the shorter EcoRI fragment, the λgt7-B10 (*b* haplotype) phage is identical to the λgt7-B10.A (*k* haplotype) phage. The greatest number of polymorphisms is

[1] In this paper, when we refer to the haplotype of a mouse strain, we are only considering the I region.

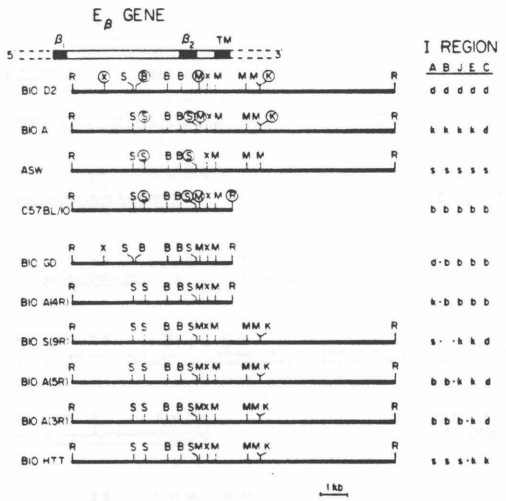


Fig. 2. Partial restriction enzyme maps of the cloned EcoRI fragments spanning the region of recombination in I-region recombinants. The upper four maps are of the parental haplotypes, *d* (B10.D2), *k* (B10.A), *s* (ASW), and *b* (C57BL/10). The lower six maps are those of I-region recombinants. The β1, β2, and transmembrane-cytoplasmic (TM) exons of the E_β gene are indicated by black boxes and are properly aligned with the restriction maps below them. R, EcoRI; X, XbaI; S, SmaI; B, BamHI; M, MstII; K, KpnI. Polymorphic restriction sites are circled. The locations of the recombination events as determined by recombinational analyses using serological markers are indicated to the right of the restriction maps

observed between the λgt7-B10.D2 (*d* haplotype) and λgt7-B10.A DNAs. The λgt7-ASW phage lacks the KpnI site found in the λgt7-B10.A phage and an MstII site found in the other three parental haplotypes. The restriction maps of the E_β gene from BALB/c (*d* haplotype) and λgt7-B10.D2 phage (*d* haplotype) are identical, suggesting that cloning artifacts that might alter restriction sites are very rare.

Mapping of the Recombination Region in the Cloned I-region Recombinants

Restriction enzyme maps were determined for the E_β gene region of the six cloned I-region recombinants (Fig. 2). The restriction enzyme sites in the DNA from the recombinants could be assigned to I subregions on the basis of immunogenetic information and restriction enzyme maps for the parental strains. B10.GD, a recombinant between the *d* and *b* haplotypes, defines the rightmost boundary of the I-A subregion by its lack of a SmaI site present in the *b* haplotype. In a similar manner, B10.HTT and B10.S(9R), recombinants between the *s* and *k* haplotypes, define the leftmost boundary of the I-E subregion by the presence of a MstII site absent in the *s* parent (Fig. 3). This result therefore maps the I-J subregion to the right of the recombination event in B10.GD and to the left of the recombination event in B10.HTT. If the I-B subregion exists, it must also lie within this region. The fact that only a single polymorphic restriction site (EcoRI) identified in the E_β gene region distinguishes the *k* from the *b* parental haplotype prevents the localization of the recombination region in B10.A(3R) and B10.A(5R) without the above information defining the subregions. At this level of analysis, B10.A(3R) and B10.A(5R) are identical. All the recombinants may fall within the 2.0-kb region between the SmaI and MstII sites in the B10.A outlined in Fig. 3. Thus, using B10.GD and B10.HTT, the maximum distance between the I-A and I-E subregions is 2.0 kb. The I-A and I-E subregions conceivably could be directly adjacent to one another, as restriction enzyme mapping can define only maximum and not minimum distances between subregions.

Localization of the Recombination Region Within the E_β Gene.

The 5′ and 3′ ends of the E_β gene were defined by hybridization to the 5′ and 3′ probes of the DC-β gene, the human homologue of the A_β gene which also shows hybridization to the E_β gene [17]. To characterize the exons of the E_β gene further, the DNA of this region was cleared into small DNA fragments (<1 kb) with restriction enzymes and hybridized to the DC-β 5′ and 3′ probes. The β1, β2, and transmembrane-cytoplasmic exons were localized to DNA fragments smaller than 600 base pairs (bp) each and are shown in Fig. 2. These data indicate that the 2.0-kb region between the I-A and I-E subregions contains the 3′ end of the large intron between the β1 and β2 exons and the β2 exon of the E_β gene.

The other striking implication of these data is that all six I region recombinants studied may map within 2 kb of one another. The distance separating I-A and I-E may in fact be slightly larger than 2.0 kb if the B10.A(4R) crossover point maps to the left of the crossover point in B10.GD and/or if the crossover point in B10.A(3R) maps to the right of the crossover point in B10.HTT. However, regardless of the precise size of this region, our data and those of Steinmetz et al. [4] clearly indicate that recombination is not random in this portion of the I region.

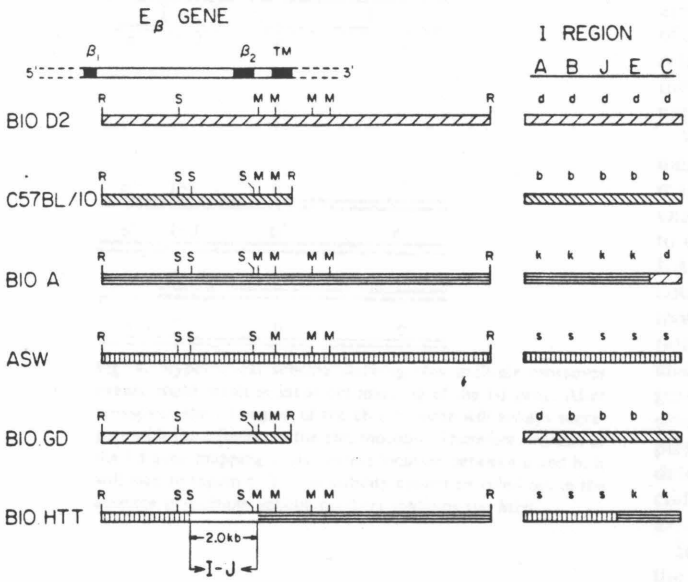


Fig. 3. Simplified restriction maps of the cloned recombination region from parental and two I-region recombinants. Parental maps have been assigned patterns that are used in the recombinant maps to indicate the parental origins of the DNA. The colinearity of the mouse genome in this region, determined on the basis of the alignment of the restriction maps from the ten different mouse strains examined, permits us to designate the 2.0-kb region within which the I-J gene should lie if the genetic mapping is correct. The 2.0-kb region separating the I-A from the I-E subregion is defined by a SmaI site absent in the recombinant B10.GD strain but present in the parental C57BL/10 strain, marking the right-hand boundary of the I-A subregion, and by a MstII site present in the recombinant B10.HTT strain but absent in the parental ASW strain, marking the left-hand boundary of the I-E subregion. The I-J gene should map to the right of the *d* haplotype DNA in B10.GD and to the left of the *k* haplotype DNA in B10.HTT. The recombination crossover point can be localized precisely in B10.GD, as indicated by the parental patterns. The left-hand boundary designation of the recombination region in B10.HTT is based on the assignment of the I-A boundary in B10.GD

Interpretation

Serological and genetic analyses of mice recombinant in the I region suggest that the I-J subregion lies between the I-A and I-E subregions. We examined the chromosomal structure of this region directly by isolating DNA fragments from the parental and recombinant mice spanning this region. Restriction enzyme site polymorphisms permitted us to localize the right-hand boundary of the I-A and the left-hand boundary of the I-E subregions to a chromosomal region approximately 2.0 kb in length. Moreover, this 2.0-kb region falls almost entirely within the second intron of the E_β gene.

Let us consider the possibility that the I-J polypeptide (serological determinant) is encoded between the I-A and I-E subregions. This distance appears to be small for a eukaryotic gene, but in fact with a few very small introns a 25,000-dalton polypeptide could be encoded on 2.0 kb of DNA, perhaps on the alternative DNA strand to that which encodes the E_β gene. Another possibility is that only one exon (encoding the I-J serological determinant) of the J gene is encoded within this region. This putative exon could be joined to the remainder of the putative I-J exons by RNA splicing or DNA rearrangements. Because of the precision of mapping by restriction enzyme site polymorphisms, we can conclude that the I-J gene product is not a result of posttranscriptional or posttranslational modifications of the full-sized E_β gene.

To test these possibilities, DNA clones encompassing the 2.0-kb region were used in Northern blots against the RNAs from 13 suppressor T cell lines (eight cells tested as positive for the I-J determinant on the cell surface) [18]. These experiments demonstrated to a precision of three RNA molecules per cell that the suppressor T cell lines do not have transcripts containing the 2.0-kb region. In addition, clones surrounding the E_β gene by 60 kb in the 5′ and 100 kb in the 3′ direction also fail to hybridize with these RNAs. Hence, RNA splicing of an I-J exon in this region appears unlikely. Moreover, the DNAs from two suppressor T cells have been examined carefully and they showed no DNA rearrangements in the E_β region. These experiments argue that the I-J polypeptide must not be encoded between the I-A and I-E subregions of the I region.

Does the I-J polypeptide exist? Several laboratories have reported the production of monoclonal antibodies for this determinant. One group has successfully immunoprecipitated this polypeptide [T. Tada, personal communication]. If the I-J polypeptide is not encoded between the I-A and I-E subregions, it must be encoded elsewhere—in the I region, in the MHC, or on another chromosome. That

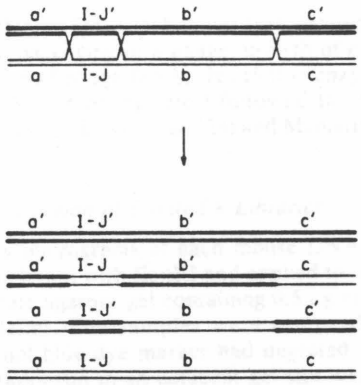


Fig. 4. Hypothetical scheme showing how multiple crossover events could result in incorrect mapping of the I-J gene. After crossover, the I-J region of the chromosome will always segregate with the c region of the chromosome. Therefore, instead of the I-J gene mapping to its correct location between a and b, it will map to region c. This possibility cannot be ruled out in the absence of multiple genetic markers spanning the MHC

the I-J gene lies elsewhere in the I region or even the MHC is reasonable only if all of the recombinant chromosomes had three (or multiples thereof) identical recombination events (Fig. 4). The nine I-region recombinants that have been studied to date all map within 8 kb of one another [4, 19]. Hence, recombination in the I region is not random. If two additional localized hotspots of recombination exist, one could posit that the I-J polypeptide is encoded elsewhere in the MHC.

A final possibility is that the I-J polypeptide is encoded elsewhere in the genome. This possibility is difficult to explain in terms of the genetic and serological data on the I-J serological determinants, but it must be considered seriously. One such explanation would be that the 2.0-kb region encodes a regulatory element that controls the expression of I-J structural genes located elsewhere. This hypothesis is unattractive because it does not explain the polymorphic nature of the I-J determinants.

The hotspot for recombination within the I region may correlate with a restriction enzyme polymorphism breakpoint in the I region [19]. DNA sequences proximal to the E_α gene are quite polymorphic compared with distal sequences. This observation also correlates with the known serological polymorphisms of the class II gene products. The known polymorphism of E_β may be related to this hotspot of recombination. The possibility that this hotspot of recombination may function to promote and ensure diversity in the MHC is worth considering. Combining alleles at K and I-A with distinct alleles at I-E and D,L could generate diversity among sets of class I genes linked to particular sets of class II genes. Additionally, these data are clearly relevant to the observation by Baxevanis et al. that the I-A and I-E subregions interact, thereby eliminating the need for the I-B subregion [7].

We plan to subclone into a plasmid vector fragments of DNA spanning the 2.0-kb region and to map the region more extensively with restriction enzymes. The resulting information may allow us to determine more precisely the boundaries of the I-A and I-E subregions. Indeed, these boundaries could be identical. In addition, we hope to define more precisely the points at which the recombination events occurred. The DNA of B10.A(3R), I-J^b, and B10.A(5R), I-J^k, will be sequenced in this region. Those sequences may enable us to identify an open reading frame for a putative partial or complete coding sequence for an I-J polypeptide or to determine unequivocally that no additional non-E_β coding sequences exist in the putative I-J gene region.

In summary, our molecular studies suggest that the I-J polypeptide cannot be encoded between the I-A and I-E subregions. Indeed, Northern blot analysis of I-J^+ suppressor T cell lines fails to detect the expression of mRNA homologous to this DNA region [18]. Similar molecular analyses of the class II genes encoding A_α, A_β, and E_α polypeptides confirm the map locations within the I-A (A_α and A_β) and I-E (E_α) subregions that were determined immunogenetically [4, 19]. Our present studies establish that the E_β gene maps within both the I-A and the I-E subregion. In conclusion, the location of the I-J gene is uncertain; this raises intriguing questions about the assumptions on which most mammalian gene mapping is based.

Materials and Methods

Materials

Mouse strains were described previously [4]. The lambda phage vector λgt7lac5 [20] was obtained from S. Scherer. Restriction enzymes, *E. coli* DNA polymerase I, and T4 ligase were purchased from New England Biolabs, Bethesda Research Laboratories, and Boehringer Mannheim. All other materials were as previously described by this laboratory [21].

Recombinant DNA Techniques

Mouse DNAs were prepared according to the method of Blin and Stafford [22]. All standard re-

combinant DNA techniques (gel electrophoresis, in vitro packaging of λ phage, growth of phage, preparation of phage DNAs, restriction enzyme digests, Southern blotting, etc.) followed the procedures outlined by Davis et al. [23] and Maniatis et al. [24].

Construction of Limited λ Libraries

Fifty micrograms of each mouse DNA was cut to completion with EcoRI and applied to a 0.7% Tris–acetate agarose gel containing 0.5 μg ethidium bromide/ml. The samples were run until the bromphenol blue dye marker had migrated 10 cm. Size markers run in an adjacent gel slot were cut away from the main gel and photographed next to a ruler. A 0.5-cm-wide region containing the DNA fragment of the desired size (either 6 kb or 12 kb) was eluted into DE-81 filter paper using a modification of the technique of Dretzen et al. [25]. The total amount of DNA recovered was approximately 1 μg. This insert DNA was mixed with an equimolar amount of EcoRI-produced arms of λgt71ac5 DNA isolated from a 0.55% low-melting agarose gel. The final DNA concentration was 200–300 μg/ml. The λ arms were annealed by heating for 10 min at 50°C and then cooled to room temperature. This step is necessary for efficient in vitro λ packaging. The ligation was carried out at 15°C overnight. The recombinant phage were in vitro packaged and plated on *E. coli* C600 [20]. These λgt7 recombinant libraries were screened with probe 4, cosmid 39.1 walking probe described by Steinmetz et al. [4]. Positively hybridizing plaques were purified and grown. DNA was prepared and used for restriction-enzyme mapping.

Acknowledgments. We thank Karyl Minard for preparation of the mouse DNAs, Elizabeth Gibb for assisting in the preparation of the λgt7-B10.A(5R) library, Connie Katz for preparation of the manuscript, and members of the Hood group for critical comments. J. K and J. M. are postdoctoral fellows of the Arthritis Foundation. This work was supported by grants from the NIH.

References

1. J. Klein (1975): Biology of the mouse histocompatibility-2 complex. Springer-Verlag, New York.
2. J. Klein, A. Juretic, C. N. Baxevanis, Z. A. Nagy (1981): The traditional and a new version of the mouse H-2 complex. *Nature 291*:455.
3. D. B. Murphy (1981): Genetic fine structure of the H-2 gene complex. In: The role of the major histocompatibility complex in immunobiology. M. E. Dorf, ed., Garland STPM Press, New York, 1.
4. M. Steinmetz, K. Minard, S. Horvath, J. McNicholas, J. Frelinger, C. Wake, E. Long, B. Mach, L. Hood (1982): A molecular map of the immune response region from the major histocompatibility complex of the mouse. *Nature 300*:35.
5. D. C. Shreffler, C. S. David (1974): The H-2 major histocompatibility complex and the I immune response region: genetic variation, function and organization. *Adv. Immunol. 20*:125.
6. J. A. Berzofsky, D. S. Pisetsky, D. J. Killion, G. Hicks, D. H. Sachs (1981): Ir genes of different high responder haplotypes for staphylococcal nuclease are not allelic. *J. Immunol. 127*:2453.
7. C. N. Baxevanis, Z. A. Nagy, J. Klein (1981): A novel type of T–T cell interaction removes the requirement for I-B region in the H-2 complex. *Proc. Natl. Acad. Sci. USA 78*:3809.
8. D. B. Murphy, L. A. Herzenberg, K. Okumura, L. A. Herzenberg, H. O. McDevitt (1976): A new I subregion (I-J) marked by a locus (Ia-4) controlling surface determinants on suppressor T lymphocytes. *J. Exp. Med. 144*:699.
9. C. Waltenbaugh (1981): Regulation of immune responses by I-J gene products. I. Production and characterization of anti-I-J monoclonal antibodies. *J. Exp. Med. 154*:1570.
10. M. Kanno, S. Kobayashi, T. Takuhisa, I. Takei, N. Shinohara, M. Taniguchi (1981): Monoclonal antibodies that recognize the product controlled by a gene in the I-J subregion of the mouse H-2 complex. *J. Exp. Med. 154*:1290.
11. T. Tada, G. Suzuki, K. Hiramatsu (1982): Some comments on the antigen receptors expressed on functional T-cell hybridomas. In: *Lymphokines. Monoclonal T cells and their products*, vol. 5. M. Feldmann, M. H. Schrier, eds., Academic Press, New York, 119.
12. T. Tada, M. Taniguchi, C. S. David (1976): Properties of the antigen-specific suppressive T-cell factor in the regulation of antibody response of the mouse. IV. Special subregion assignment of the gene(s) that codes for the suppressive T-cell factor in the H-2 histocompatibility complex. *J. Exp. Med. 144*:713.
13. D. B. Murphy (1978): The I-J subregion of the murine H-2 gene complex. *Springer Semin. Immunopathol. 1*:111.
14. M. Taniguchi, T. Tokuhisa, M. Kanno, Y. Yaoita, A. Shimizu, T. Honjo (1982): Reconstitution of antigen-specific suppressor activity with translation products of mRNA. *Nature 298*:172.
15. J. Trial, J. A. Kapp, C. W. Pierce, D. C. Shreffler, C. M. Sorensen, C. Waltenbaugh (1983): Expression of cell surface antigens by suppressor T cell hybridomas. I. Comparison of phenotype and function. *J. Immunol. 130*:565.
16. K. Krupen, B. A. Araneo, L. Brink, J. A. Kapp, S. Stein, K. J. Wieder, D. R. Webb (1982): Purification and characterization of a monoclonal T-cell suppressor factor specific for poly ($(LGlu^{60}LAla^{30}LTyr^{10})$). *Proc. Natl. Acad. Sci. USA 79*:1254.
17. E. O. Long, C. T. Wake, M. Strubin, N. Gross, R. S. Accolla, S. Carrel, B. Mach (1982): Isolation of distinct cDNA clones encoding HLA-DR β-chains by use of an expression assay. *Proc. Natl. Acad. Sci. USA 79*:7465.
18. M. Kronenberg, M. Steinmetz, J. Kobori, E. Kraig, J. A. Kapp, C. W. Pierce, G. Suzuki, T. Tada, L. Hood (1983): RNA transcripts for I-J polypeptides are apparently not encoded between the I-A and I-E subregions of the murine major histocompatibility complex. *Proc. Natl. Acad. Sci. USA 80*:5704.
19. L. Hood, M. Steinmetz, B. Malissen (1983): Genes of the

major histocompatibility complex of the mouse. *Annu. Rev. Immunol. 1*:529.
20. L. Rowen, J. Kobori, S. Scherer (1982): Cloning of bacterial DNA replication genes in bacteriophage λ. *Mol. Gen. Genet. 187*:501.
21. M. Steinmetz, A. Winoto, K. Minard, L. Hood (1982): Cluster of genes encoding mouse transplantation antigens. *Cell 28*:489.
22. N. Blin, D. W. Stafford (1976): A general method for isolation of high molecular weight DNA from eukaryotes. *Nucleic Acids Res. 3*:2303.
23. R. W. Davis, D. Botstein, J. R. Roth (1980): Advanced bacterial genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
24. T. Maniatis, E. F. Fritsch, J. Sambrook (1982): Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
25. G. Dretzen, M. Bellard, P. Sassone-Corsi, P. Chambon (1981): A reliable method for the recovery of DNA fragments from agarose and acrylamide gels. *Anal. Biochem. 112*:295.

Received June 27, 1983

A Critique

The authors state that the crossovers in the six recombinant strains examined occurred within a 2-kb stretch of DNA. This conclusion is not warranted from the data presented. Clearly, the crossover in B10.GD is situated in a 2-kb segment defined by the first BamHI site and the second SmaI site. However, the crossovers in B10.A(4R), B10.A(5R), and B10.A(3R) occurred somewhere between the β1 exon and the second EcoRI site (progenitor strains B10 and B10.A are identical for this segment of DNA), while the crossovers in strains B10.HTT and B10.S(9R) occurred between the β1 exon and the first MstII site. Thus, although the crossovers in all six strains occurred within the E_β gene, the crossover in B10.GD can be localized to a 2-kb segment of DNA, but the crossovers in the five other recombinants can not.

I am also concerned about the conclusion that the I-A and I-E subregions are separated by no more than 2 kb of DNA. The right-hand (D-end) boundary of the I-A subregion was defined by the crossover in strain B10.A(4R), not that in strain B10.GD. Previously, it was shown that crossovers in both strains separated marker loci mapping in the I-A subregion from other I-region marker loci. However, whether the crossovers in both strains occurred at the same site or at different sites could not be determined. Thus, from a historical viewpoint, the distance between the I-A and I-E subregions should be estimated on the basis of the distance between the crossovers in strains B10.A(4R) and B10.HTT. The traditional concepts of regions and subregions will, however, probably become obsolete when the MHC is fully defined at the molecular level.

The Authors Respond

The reviewer's comments were very helpful. We have added to the text to emphasize the limitations of our ability to analyze both the I-A and the I-E boundaries in each recombinant. The lack of additional haplotype-specific restriction site polymorphisms between the *k* and *b* haplotypes prevents us from doing the desired analysis.

We used B10.GD to define the I-A boundary because we were unable to find an appropriate restriction site polymorphism to analyze B10.A(4R). We realize that B10.A(4R) is the preferred recombinant for analysis. However, we included Table 1 (adapted from reference 3) to indicate the operational definitions of the subregions. It is difficult to extrapolate satisfactorily a "genetic" definition of the boundaries of the subregions to a "molecular" one. Regardless of which recombinant mouse strain is used to define the I-A (B10.A(4R) or B10.GD) or I-E (B10.A(5R) or B10.HTT) boundary, our conclusion of the I-J gene location as immunogenetically mapped is indisputable. The I-J gene maps to the left of the I-E subregion and to the right of the I-A subregion as mapped by restriction enzyme site polymorphisms.

APPENDIX C

STUDY ON THE PLATELET DERIVED GROWTH FACTOR GENE: THE GENOMIC COSMID CLONES OF HUMAN AND MOUSE PDGF GENE

Project

To isolate the genomic clones of human and mouse platelet derived growth factor (PDGF) gene and to establish the exon intron structure of the complete gene.

Results

I have constructed a human cosmid library from the osteosarcoma cell line U2-OS in the cosmid vector pTL5. The library was screened with the nick translated v-sis DNA. Three positive clones were obtained (cosmid 56.1, 49.1, and 44.2). The cosmid 56.1 contains the complete PDGF gene (c-sis) by several criteria: 1. Restriction enzyme mapping and hybridization to the v-sis probe indicate that the cosmid 56.1 contains all five exons hybridizing to the v-sis DNA. Hybridization to the mouse PDGF gene (see below) revealed an additional exon at the 5' end of the cosmid that is conserved between mouse and human (Figure 1). 2. Subsequent sequencing works indicated that six exons of the PDGF gene encodes the complete protein and there is no change in the nucleotide sequence of this human osteosarcoma PDGF gene from the germline PDGF gene (Arlen Thomasson, personal communication). 3. Functional PDGF protein is produced when the cosmid 56.1 is introduced into the cos cell line (Arlen Thomasson, personal communication).

To establish the exon intron structure of the PDGF gene, I have isolated genomic clones encoding the mouse PDGF gene from two cosmid libraries. They encompass 71 kb of DNA (Figure 2, clones c-sis 26 and c-sis 30 are from the BALB/c library ; clones c-sis 1,4,5,6,8 and 9 are from B10.dm1 cosmid library), and a cross hybridization study between the mouse and human PDGF gene has revealed an additional PDGF exon that is conserved evolutionarily.

In conclusion, I have isolated 30 kb of human PDGF genomic DNA and 71 kb of mouse PDGF genomic DNA; a cross hybridization study between human and

mouse clones revealed an additional PDGF exon, apart from the five exons that hybridizes to the v-sis DNA. Subsequent works by Arlen Thomasson at AMGen confirmed that the human PDGF gene contains six exons/ The isolated human cosmid clone 56.1 encompasses the complete PDGF gene, since transfection of this cosmid clone into the cos cell produces functional PDGF protein.

Figure 1. Human osteosarcoma U2OS PDGF gene. Open boxes represent the cosmid vector pTL5. Shaded box represents the new exon conserved between human and mouse. Solid box represents the five exons hybridized to the v-sis probe. The cosmid is 5' to 3' according to the transcriptional orientation.

HUMAN c-sis (PDGF) GENE FROM THE OSTEOSARCOMA U-2 CELL LINE

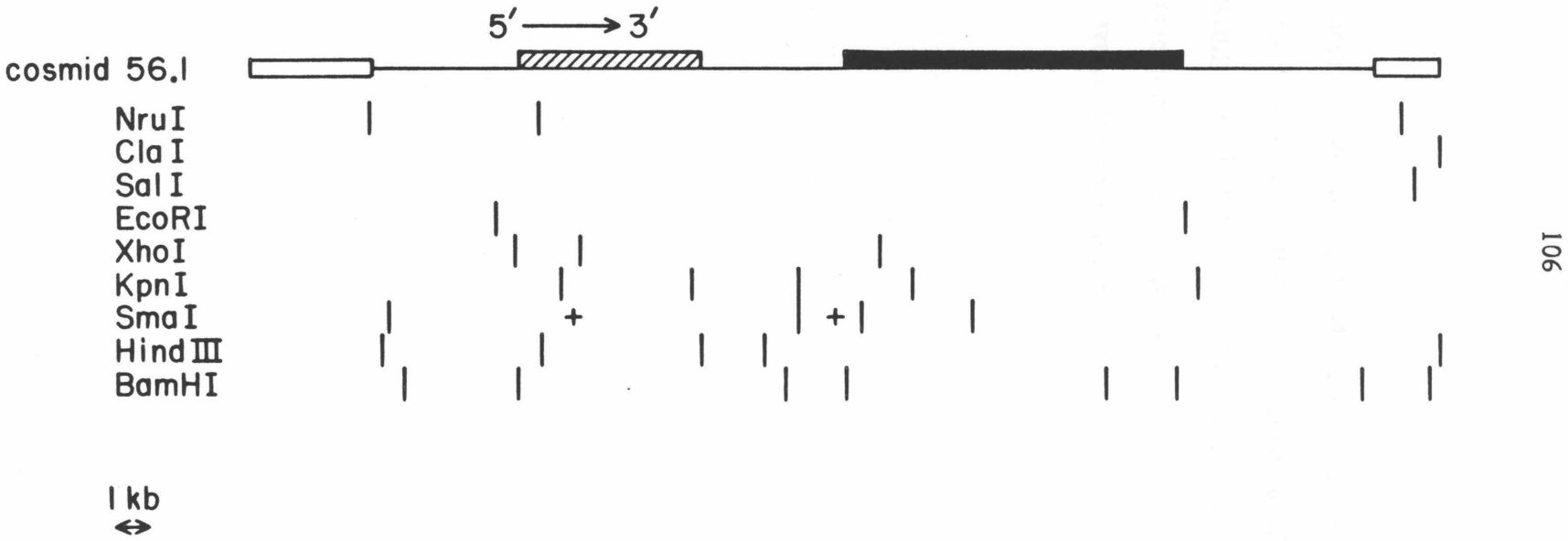
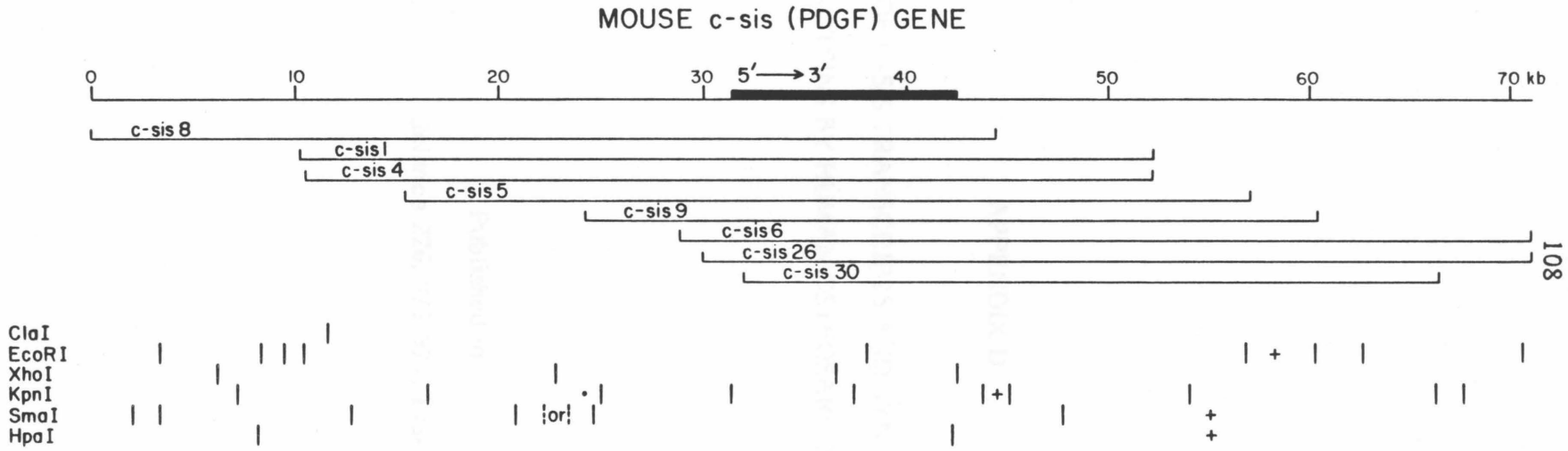


Figure 2. Cosmid clones encompassing 71 kb of the mouse PDGF gene. Cosmid clones c-sis 26 and c-sis 30 are from a Balb/c liver genomic library. The remaining clones are from a B10.D2-H-2^{dm1} library. The solid box represents the mouse PDGF exons that hybridized to the v-sis probe. The cosmid clones are oriented 5' to 3' according to the transcriptional direction.


MOUSE c-sis (PDGF) GENE
0
10
20
30
5′ → 3′
40
50
60
70 kb
c-sis 8
c-sis 1
c-sis 4
c-sis 5
c-sis 9
c-sis 6
c-sis 26
c-sis 30
108
ClaI
EcoRI
XhoI
KpnI
SmaI
HpaI
or

APPENDIX D

DETECTION OF C-SIS TRANSCRIPTS AND SYNTHESIS OF PDGF-LIKE PROTEINS BY HUMAN OSTEOSARCOMA CELLS

Published in

Science **226,** 972-974, 1984

Reprint Series
23 November 1984, Volume 226, pp. 972–974

SCIENCE

Detection of c-*sis* Transcripts and Synthesis of PDGF-Like Proteins by Human Osteosarcoma Cells

D. T. Graves, A. J. Owen, R. K. Barth, P. Tempst, A. Winoto, L. Fors, L. E. Hood, and H. N. Antoniades

Detection of c-*sis* Transcripts and Synthesis of PDGF-Like Proteins by Human Osteosarcoma Cells

Abstract. *Platelet-derived growth factor (PDGF) has been previously shown to be homologous to the transforming gene of simian sarcoma virus (v-*sis*), and inappropriate expression of the cellular counterpart of the v-*sis* gene (c-*sis*) has been implicated in the generation of mesenchymal tumors. The U-2 OS human osteosarcoma line was shown to contain multiple c-*sis* transcripts. Immunoprecipitation experiments with antiserum to PDGF identified a variety of polypeptides ranging in size from 18,000 to 165,000 daltons that were immunoprecipitated specifically from U-2 OS cell extracts. The osteosarcoma also was shown to secrete a 29,000-dalton protein having the serological and structural characteristics of PDGF.*

A variety of transformed cells have been reported to produce mitogenic factors (*1*), suggesting that the ability to produce such factors is important in the establishment and maintenance of transformation (*2*). The human osteosarcoma-derived cell line U-2 OS elaborates a mitogen similar to platelet-derived growth factor (PDGF) (*3*) and does not require exogenous PDGF for growth in vitro (*4*). PDGF represents the major mitogen of human serum for cells of mesenchymal origin (*5*). It is stored in the α-granules of platelets (*6*) and has been purified from human serum (*7*), clinically outdated blood platelets (*8*), and platelet-rich plasma (*9*). Elucidation of its amino-terminal amino acid sequence suggested that it consists of two homologous polypeptide chains (PDGF-1 and PDGF-2) linked by disulfide bonds (*10*). PDGF shares extensive sequence homology with $p28^{sis}$, the oncogene product of the simian sarcoma virus (SSV), an acute transforming primate retrovirus (*11*). In SSV-transformed cells, $p28^{sis}$ is processed into a disulfide-linked dimer structurally similar to the PDGF dimer, which is secreted by the cells and is specifically immunoprecipitated with antiserum to PDGF (*12*). In addition, a number of cell lines derived from mesenchymal tumors have been reported to contain RNA (c-*sis*), which hybridizes to a v-*sis* DNA probe (*13*).

We used a DNA probe to v-*sis* and antiserum to PDGF to investigate the synthesis of PDGF-like products by U-2 OS cells. To determine whether c-*sis* transcripts are synthesized in U-2 OS cells, we isolated total polyadenylated [poly(A)$^+$] RNA from the cells and subjected it to RNA blot analysis with a ^{32}P-labeled v-*sis* DNA probe (*14*). Two bands corresponding to RNA species of 4.0 and 3.7 kilobases (kb) in length were detected (Fig. 1A). Occasionally a smaller, more diffuse band (~2.6 kb) was also observed. We believe that the 4.0-kb RNA corresponds to the 4.2-kb c-*sis* RNA species identified earlier in a number of human sarcomas and glioblastomas (*13*). Comparison of the intensity of each c-*sis* RNA band to a dilution series of a standard RNA indicated that each cell contains 10 to 20 copies of the c-*sis* transcript (data not shown). The different sizes of the c-*sis* transcripts may be due to differences in the initiation or termination sites of RNA synthesis, differential RNA splicing, or the expression of different but related genes, or they may reflect somatic mutation of one or both alleles. The size of the transcript and the lack of integrated proviral DNA in the U-2 OS genome (data not shown) make it unlikely that the hybridizing species are derived from an SSV v-*sis* gene. No c-*sis* RNA was detected in total poly(A)$^+$ RNA isolated from MG-63 osteosarcoma cells (Fig. 1B), which do not produce PDGF (*15*).

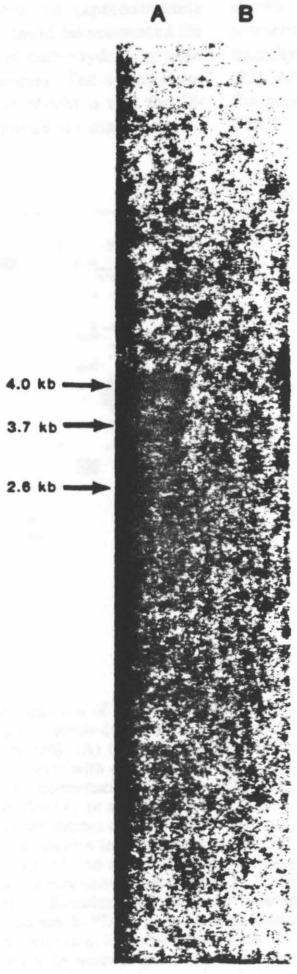


Fig. 1. RNA blot analysis of poly(A)$^+$ RNA from U-2 OS and MG-63 human osteosarcoma cells with the use of a v-*sis* DNA probe (*14*). Poly(A)$^+$ RNA was isolated by extraction in 5*M* guanidinium thiocyanate and subsequent oligo(dT) cellulose chromatography. Poly(A)$^+$ RNA (5 μg) was subjected to electrophoresis on a 1.2 percent agarose gel and transferred to nitrocellulose. The filter was then hybridized with a nick-translated Pst I restriction fragment of simian sarcoma proviral DNA, which contains the v-*sis* region. (Lane A) U-2 OS; (lane B) MG-63.

The 4.0- and 3.7-kb RNA species have a potential coding capacity of 130 to 150 kilodaltons (kD), which is approximately four times the molecular size of mature PDGF (*10*). Thus, it is possible that the mature form of the PDGF-like factor is derived from a very large precursor. To investigate these possibilities, we labeled U-2 OS cells with [^{35}S]cysteine, immunoprecipitated cytoplasmic extracts with antiserum to purified PDGF (*16*), and analyzed the precipitate by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) and autoradiography (*17*). Proteins with molecular sizes of 56, 34, 32, and 18 to 20 kD were detected with antiserum to PDGF but not with control serum on a 16 percent polyacrylamide gel run under nonreducing conditions (Fig. 2A, lanes 1 and 2). This experiment was repeated with 8 percent polyacrylamide gels to detect the large forms (Fig. 2A, lanes 3 to 5). Proteins with molecular sizes of 220, 210, and 54 to 60 kD were precipitated by antiserum to PDGF but not in the presence of excess PDGF (1 μg) or with control serum.

Under reducing conditions, the amounts of 210- to 220-kD species were much lower. Instead, a prominent 165-kD protein was precipitated in addition to minor species of 130, 115, and 90 kD. Precipitated material migrating in the 56- to 62-kD range was also detected and probably corresponded to the 54- to 60-kD protein identified under nonreducing conditions. Precipitation of all these species was inhibited by the addition of excess PDGF. The apparent shift in the mobility of the reduced 165-kD protein to proteins of 210- to 220-kD under non-

reducing conditions suggests that it is covalently associated with one or more other proteins in its native state.

Experiments were undertaken to analyze the forms of PDGF-like polypeptides secreted by U-2 OS cells. Cysteine-free minimum essential medium (MEM) was conditioned by U-2 OS cells in the presence of [^{35}S]cysteine (*18*), concentrated, immunoprecipitated as described (*17*), and analyzed on a 16 percent gel (Fig. 2B). Under nonreducing conditions, a protein with a molecular size of approximately 200 kD and one of 29 kD were precipitated by antiserum to PDGF but not by control serum or in the presence of excess PDGF (Fig. 2B, lanes 1 to 3). Reduced precipitates had molecular sizes of 110, 16, and 15 kD (lanes 4 to 6). The 200- and 29-kD proteins present under nonreducing conditions disappeared upon reduction, suggesting that they were derived from the 110-, 16-, and 15-kD proteins, respectively. A 29-kd protein that can be reduced to a 16- and 15-kD polypeptide is consistent with the structure of PDGF.

To establish the specificity of the immunoprecipitation assay, we repeated the immunoprecipitation experiments with isotopically labeled MG-63 cells and GM-10 human diploid fibroblasts (Fig. 2C). No difference was detected when the precipitation was carried out with antiserum to PDGF or control serum.

That antiserum to PDGF cross-reacts with biologically active factors secreted by U-2 OS cells was confirmed by partially purifying U-2 OS medium as described (*19*) and assaying it for stimulation of [^{3}H]thymidine incorporation in 3T3 cells (Table 1). The amount of [^{3}H]thymidine incorporated into 3T3 cells in partially purified conditioned medium was five times greater than that in control. Antiserum to PDGF blocked 81 percent of the mitogenic activity of this conditioned medium.

The results indicate that the mitogenic factor secreted by U-2 OS is PDGF. First, U-2 OS cells were shown to produce RNA transcripts that are homologous to the c-*sis* proto-oncogene. The c-*sis* gene has recently been shown to encode the PDGF-2 chain (*20*). Also, DNA sequence analysis of complementary DNA (cDNA) clones constructed from U-2 OS poly(A)$^+$ RNA confirms that the osteosarcoma contains transcripts derived from the c-*sis* gene (*21*). Second, immunoprecipitation of conditioned medium from U-2 OS cells revealed that they secrete a 29-kD protein that shares antigenic determinants with PDGF. This 29-kD protein is a disulfide-linked dimer of 15- and 16-kD proteins,

Table 1. Effect of antiserum to PDGF on partially purified osteosarcoma-derived growth factor.

Sample	[^{3}H]Thymidine incorporation (count/min)	Inhibition (percent)
CM-Sephadex preparation	127,561	
+ antiserum to PDGF (1:1000)	84,155	34
+ antiserum to PDGF (1:333)	24,028	81
Control	19,462	

a structural feature characteristic of PDGF. Third, the same antiserum used in the experiments described above blocked the mitogenic activity of medium conditioned with U-2 OS.

Precipitation of U-2 OS cell extracts revealed several cross-reacting proteins ranging from 18 to 20 kD up to 165 kD. The smallest proteins (18 to 20 kD) have an apparent size that agrees well with the size of the reduced PDGF chains. The 165-kD protein was slightly larger than the total coding capacity of the largest c-*sis* transcript detected (approximately 150 kD), but this could be accounted for by the presence of carbohydrate groups on the 165-kD species. The simplest interpretation of our results is that the 18- to 20-kD polypeptides are mature forms of PDGF chains, the 165-kD polypeptide is a PDGF precursor, and the remaining species represent processing intermediates. We have, however, no direct proof that this interpretation is correct, and alternative explanations are possible. For example, one or more of the polypeptides we have identified may not have PDGF antigenic determinants but might be tightly associated with peptides that do react with antiserum to PDGF and, consequently, are precipitated together with them. Recently, it has been shown that cell extracts treated with antiserum to p21ras coprecipitate the transferrin receptor with p21ras protein (*22*). Alternatively, the different sizes of the proteins reactive with antiserum to PDGF may reflect the multiple RNA

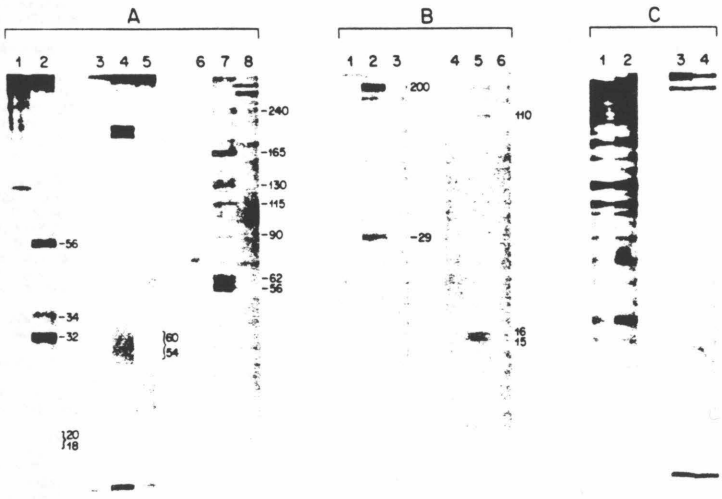


Fig. 2. Immunoprecipitation of PDGF-like peptides from U-2 OS human osteosarcoma cells. Numbers to the right of bands denote kilodaltons. U-2 OS cells were labeled and immunoprecipitated as described (*16*). (A) Cell extracts analyzed on 16 percent gels under nonreducing conditions after treatment with normal rabbit serum (lane 1) or antiserum to PDGF (lane 2); on an 8 percent gel under nonreducing conditions after treatment with normal rabbit serum (lane 3), antiserum to PDGF (lane 4), or antiserum to PDGF plus excess PDGF (lane 5); and on 8 percent gels under reducing conditions after treatment with normal rabbit serum (lane 6), antiserum to PDGF (lane 7), or antiserum to PDGF plus excess PDGF (lane 8). (B) Labeled medium was prepared as described (*18*) and analyzed on 16 percent gels under nonreducing conditions after exposure to normal serum (lane 1), antiserum to PDGF (lane 2), or antiserum to PDGF plus excess PDGF (lane 3). Samples analyzed under reducing conditions were treated with normal serum (lane 4), antiserum to PDGF (lane 5), or antiserum to PDGF plus excess PDGF (lane 6). (C) Labeled MG-63 cell extracts treated with normal serum (lane 1) or antiserum to PDGF (lane 2) and analyzed on a 16 percent gel under nonreducing conditions. GM-10 cells, similarly prepared, were treated with normal serum (lane 3) or antiserum to PDGF (lane 4) and analyzed on an 8 percent gel under reducing conditions.

species found in these cells. The complete characterization of c-*sis* cDNA clones, as well as experiments designed to elucidate the precursor-product relations of the various proteins, should help to determine the relevance of these proteins to the synthesis, assembly, and function of PDGF in U-2 OS cells.

D. T. GRAVES
A. J. OWEN
Departments of Periodontology and *Nutrition, Harvard School of Public Health, Boston, Massachusetts 02115*
R. K. BARTH
P. TEMPST
A. WINOTO
L. FORS
L. E. HOOD
Division of Biology, California Institute of Technology, Pasadena 91125
H. N. ANTONIADES
Center for Blood Research and *Department of Nutrition, Harvard School of Public Health*

References and Notes

1. M. A. Anzano *et al.*, *Cancer Res.* **42**, 4776 (1982); J. E. DeLarco and G. J. Todaro, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 4001 (1978); B. Ozanne, R. J. Fulton, P. L. Kaplan, *J. Cell Physiol.* **110**, 81 (1982); A. B. Roberts *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3494 (1980).
2. H. L. Moses and R. A. Robinson, *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **41**, 3008 (1982); P. L. Kaplan, M. Anderson, B. Ozanne, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 485 (1982); M. B. Sporn, G. J. Todaro, *N. Engl. J. Med.* **303**, 878 (1980).
3. C.-H. Heldin, B. Westermark, A. Wasteson, *J. Cell Physiol.* **105**, 235 (1980); C. Betsholtz *et al.*, *Biochem. Biophys. Res. Commun.* **117**, 176 (1983).
4. D. T. Graves, A. J. Owen, H. N. Antoniades, *Cancer Res.* **43**, 83 (1983).
5. H. N. Antoniades, D. Stathakos, C. D. Scher, *Proc. Natl. Acad. Sci. U.S.A.* **70**, 675 (1973); N. Kohler and A. Lipton, *Exp. Cell Res.* **87**, 297 (1974); R. Ross, J. Glomset, B. Kariya, L. Harker, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 1207 (1974); B. Westermark, A. Wasteson, K. Uthne, *Exp. Cell Res.* **96**, 58 (1975).
6. D. R. Kaplan, F. C. Chao, C. D. Stiles, H. N. Antoniades, C. D. Scher, *Blood* **53**, 1043 (1979).
7. H. N. Antoniades, D. Stathakos, C. D. Scher, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 2635 (1975).
8. H. N. Antoniades, C. D. Scher, C. S. Stiles, *ibid.* **76**, 1809 (1979); H. N. Antoniades, *ibid.* **78**, 7314 (1981); C.-H. Heldin, B. Westermark, A. Wasteson, *ibid.* **76**, 3722 (1979); *ibid.* **78**, 3664 (1981).
9. T. F. Deuel *et al.*, *J. Biol. Chem.* **256**, 8896 (1981); E. W. Raines and R. Ross, *ibid.* **257**, 5154 (1982).
10. H. N. Antoniades and M. W. Hunkapiller, *Science* **220**, 963 (1983).
11. R. F. Doolittle *et al.*, *ibid.* **221**, 275 (1983); M. D. Waterfield *et al.*, *Nature (London)* **304**, 35 (1983).
12. K. C. Robbins, H. N. Antoniades, S. G. Devare, M. W. Hunkapiller, S. Aaronson, *Nature (London)* **305**, 605 (1983). A. J. Owen, P. Pantazis, H. N. Antoniades, *Science* **225**, 54 (1984).
13. A. Eva *et al.*, *Nature (London)* **295**, 116 (1982).
14. J. M. Chirgwin, A. E. Przybyla, R. J. MacDonald, W. J. Rutter, *Biochemistry* **18**, 5294 (1979); H. Aviv and P. Leder, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 1408 (1972); P. S. Thomas, *ibid.* **77**, 5201 (1980); P. W. J. Rigby, M. Dieckmann, C. Rhodes, P. Berg, *J. Mol. Biol.* **113**, 237 (1977); K. C. Robbins, S. G. Devare, E. P. Reddy, S. A. Aaronson, *Science* **218**, 1131 (1982).
15. D. T. Graves, S. R. Williams, H. N. Antoniades, A. J. Owen, *Cancer Res.* **44**, in press.
16. Antiserum to PDGF was produced in rabbits as described by A. J. Owen *et al.* [*Proc. Natl. Acad. Sci. U.S.A.* **79**, 3203 (1982)]. Confluent U-2 OS cells were incubated for 2 hours in cysteine-free MEM containing 300 μCi of carrier-free [^{35}S]cysteine per milliliter. The cells were collected (~ 10^7 per batch) and lysed with immunoprecipitation buffer [0.3*M* NaCl, 0.02*M* tris-HCl (*p*H 7.5), 0.001*M* EDTA, 0.01*M* NaI, 0.002*M* phenylmethylsulfonyl fluoride (PMSF), 0.3 percent Triton X-100, and 1 percent human serum albumin (HSA). The extracts were cleared with 50 μg of Protein A–Sepharose. Normal rabbit serum or antiserum to PDGF (2 μl) was then added to equal portions of the cell extract and allowed to incubate for 18 hours at 4°C. Immune complexes were recovered by the addition of Protein A–Sepharose (40 μg).
17. The washed Protein A–Sepharose beads were resuspended in SDS sample buffer [50 mM tris-HCl (*p*H 6.8), 2 percent SDS, 1 m*M* PMSF, and 12 percent glycerol] and boiled for 3 minutes. The beads were pelleted, and the supernatants were collected and divided into two equal portions. Dithiothreitol (100 m*M* final concentration) was added to one of the samples and left for 90 minutes at 37°C. 2-Mercaptoethanol (10 percent final volume) was added to the reduced sample 15 minutes before applying the sample to gels (30 by 20 cm by 0.8 mm SDS-PAGE). Resolving gels (8 percent) were used to analyze the large peptides and 16 percent gels for the small peptides. The gels were run as described [P. Pantazis and W. M. Bonner, *J. Biol. Chem.* **256**, 4669 (1981)]. Gels were dried and autoradiographed as described [W. M. Bonner, M. H. P. West, J. D. Stedman, *Eur. J. Biochem.* **109**, 17 (1980)].
18. Labeled conditioned medium from U-2 OS cells was prepared and immunoprecipitated as described by A. J. Owen *et al.* [in (*12*)].
19. CM-Sephadex that had been swollen was added to MEM conditioned with U-2 OS cells (100 ml), dialyzed overnight against water, poured on a column, and eluted with 1*M* NaCl. The eluted fractions were dialyzed exhaustively against acetic acid, lyophilized, and resuspended in HSA (1 percent).
20. S. F. Joseph, C. Guo, L. Ratner, F. Wong-Staal, *Science* **223**, 487 (1984); I.-M. Chiu *et al.*, *Cell* **37**, 123 (1984).
21. P. Tempst *et al.*, in preparation.
22. T. Finkel and G. M. Cooper, *Cell* **36**, 115 (1984).
23. We thank S. Williams, R. Siraco, G. Easterly, C. Katz, and J. MacLaren for technical assistance. Supported by National Cancer Institute grant CA30101 (H.N.A.) and NIH grants HL29583 (H.N.A.) and HL27607 (A.J.O.). R.K.B. is a fellow of the Anna Fuller Fund and P.T. is a special fellow of the Leukemia Society of America.

18 May 1984; accepted 27 July 1984

APPENDIX E

IDENTIFICATION AND EXPRESSION OF GENES ENCODING RAT CLASS I AND II MAJOR HISTOCOMPATIBILITY COMPLEX MOLECULES FROM TWO COSMID LIBRARIES.

Published in

Transplantation Proceedings 17, 1808-1811, 1985

Identification and Expression of Genes Encoding Rat Class I and II Major Histocompatibility Complex Molecules From Two Genomic Cosmid Libraries

A.G. Diamond, J.M. Windle, G.W. Butcher, A. Winoto, L. Hood, and J.C. Howard

ANALYSIS of the rat major histocompatibility complex (MHC) (*RT1*) through serologic and cell-mediated assays of inbred, congenic, and recombinant lines has produced a linkage map that shows overall two regions encoding class I genes separated by one encoding class II.[1,2] The *RT1.A* region encodes the serologically immunodominant, highly expressed class I molecule(s), which can easily be shown to function as restriction elements for T cell-mediated responses and are thus equivalents of H-2K, -D, and -L. The molecules encoded by the *RT1.C* and *RT1.E* regions show more restricted distribution, a lower level of expression, and, although not extensively characterized either serologically or functionally, thus appear to be similar to the mouse Qa/T1a antigen series. At least two class II molecules have been well characterized and mapped to the *RT1.B* and *RT1.D*, regions that encode I-A– and I-E–like molecules, respectively.[3] However, although much progress has been made, the number of molecules encoded by the various regions of *RT1* remains uncertain.

Construction of cosmid libraries and expression of cell surface products of cloned genes in L cell fibroblast lines has facilitated identification and orientation of clusters of MHC genes of other species, particularly the mouse.[4,5] We have initiated a study of the *RT1* complex by construction of cosmid libraries from PVG-*RT1*^{a} and PVG (*RT1*^{c}) animals. These have been screened using mouse DNA probes for class I and II genes. DNA encoding putative rat MHC genes has been characterized by restriction mapping, functional genes detected by introduction into mouse Ltk$^-$ fibroblasts, and expressed products identified by fluorescence-activated cell sorter (FACS) analysis with monoclonal antibodies (MAbs).

MATERIALS AND METHODS

Construction of Libraries, Screening, and Isolation of Cosmid DNA

Procedures used were essentially as described by Steinmetz et al,[6] using liver DNA obtained from PVG and PVG-*RT1*^{a} rats (bred at IAP).

Probes

The probes used are summarized in Table 1. With the exception of the cDNA clone pH-2 IIa,[6] all probes were fragments isolated from mouse cosmid libraries. Where these had been previously subcloned, the insert was excised and used. DNA was nick-translated with ^{32}P-labeled deoxynucleotide triphosphates to a specific activity of around 10^8 cpm/μg. For both screening and characterization of cosmid DNA, final washes were in 1 × saline sodium citrate (SSC), 0.1% sodium dodecyl sulfate (SDS) at 68 °C.

Gene Transfer Into Ltk$^-$ Fibroblasts

Calcium phosphate-mediated transformation of Ltk$^-$ fibroblasts was as described by Malissen et al.[8] Cosmid clones were linearized with *Cla*I or *Nru*I, and 1 μg (class I) or 6 μg (class II) was added to 1 to 2 × 10^6 cells in a 75-cm^2 flask. Ltk$^-$ carrier DNA was used to bring the total to 20 μg. Plasmids containing herpes simplex virus (HSV) tk gene, ptk5 (some class I), and pBR 322/ASV-2LTR/tkA (remainder of class I and all of class II) were linearized with *Hind*III and 50 ng was added. No difference was observed between the two plasmids. Hypoxanthine-aminopterin-thymidine (HAT)-resistant lines were grown to confluence before assaying for anti-

From the Department of Immunology, Agricultural and Food Research Council Institute of Animal Physiology (IAP), Babraham, Cambridge, England, and the Division of Biology, California Institute of Technology, Pasadena.

Supported in part by National Institutes of Health grants No. CA34913 and AI19624 and Medical Research Council grant No. 979/456/C.

Address reprint requests to Dr A.G. Diamond, Department of Immunology, AFRC Institute of Animal Physiology, Babraham, Cambridge, CB2 4AT, England.

0041-1345/85/1703-0004$03.00/0

Table 1. Probes Used for Screening and Characterization

H-2 Gene Recognized	Probe	Origin	Description	Size (kb)
Class I				
K	cos H-2^d 17.1 *KpnI/SmaI*	BALB/c cosmid library	Low-copy region 3′ of K^d	2.6
D	cos H-2^d 18.1 *NruI/SmaI*	BALB/c cosmid library	Low-copy region 5′ of D^d	3.2
L	mp 8.59.2	Subclone from BALB/c cosmic library	Low-copy region 5′ of L^d	0.8
All	pH-2IIa *HhaI/SacI*	C14 (BALB/c) lymphoma cDNA library	Third and TM domains of class I	0.4
Class II				
A_α	p24.2	Subclone from BALB/c cosmid library	A^d_α	4.8
A_β	A_βpBR	Subclone from BALB/c cosmid library	A^d_β	5.6
$A_{\beta 2}$[7]	cos 25.1[k] *ClaI/HpaI*	AKR cosmid library	$A^k_{\beta 2}$	2.0
E_α	32.1.2mp8	Subclone from BALB/c cosmid library	1, 2, and TM domains of E^d_α	3.4
E_β	pBR325-8.48	Subclone from BALB/c cosmid library	1 and 2 domains of E^d_β.	7.2

TM, transmembrane.

gen expression. Clones were established by growth at limiting dilution.

Antigen Expression Assay

Cells were removed from flasks with EDTA, washed in cold azide-containing medium, and incubated in round-bottomed microtiter trays with MAb tissue culture supernatant with azide for one hour at 4 °C. After three washes, fluorescein isothiocyanate (FITC)-labeled affinity-purified sheep anti-mouse IgG or rabbit anti-rat IgG immunoglobulin was added. Following an additional one hour at 4 °C, cells were washed and analyzed in a Becton Dickinson (Mountain View, Calif) FACS 420.

Monoclonal Antibodies

Properties and origins of the rat alloantibodies R2/15S, R3/13, YR1/100, and YR1/146 specific for separate sites on RT1.A^a, and the mouse anti-rat class I antibodies F16.4.4, MRC OX18, and 29/18 have been summarized.[9] MRC OX4 and OX6 MAb were used to detect RT1.B[10] and MRC OX17 for RT1.D.[11]

RESULTS AND DISCUSSION

Class I Genes

The PVG-*RT1*^{a} library was screened with "low-copy" probes specific for the mouse *H-2D* and *H-2L* loci.[12] A total of 43 cosmid clones were obtained. Plasmid DNA isolated from all of these hybridized with the H-2D probe, with 12 also showing hybridization to the H-2L probe. Further analysis showed that 15 of these cosmids also hybridized with a probe for the *H-2K* region (Table 2). As cosmids hybridized to two or even all three of these low-copy sequences, they cannot be used to identify specific loci in the rat that are equivalents of mouse *H-2K*, *-D*, or *-L*. Indeed, a recent study has shown that even in the mouse there is extensive cross-hybridization between similar low-copy probes and cosmids from *H-2K* and *Qa-2,3* regions.[5] Nevertheless, all of these cosmids show at least one, but more often two or three, *Eco*RI fragments that hybridize to the mouse class I cDNA probe, pH-2 IIa, implying that they contain as many class I genes. These noncoding sequence probes are, therefore, still specific for DNA flanking class I MHC genes that must have been conserved since the separation of rats and mice in evolution.

To identify the molecules encoded, DNA from each cosmid was introduced into mouse L cell fibroblasts, and transformant lines were analyzed for expression with MAb to rat class I molecules. None of the alloantibodies to RT1.A^a reacted with any transformant line. The xenoantibody MRC OX18, but not F16.4.4 or 29/18, reacted with 12 of the 43 lines. Thus, molecules were expressed that are unlikely to be RT1.A (although an influence

Table 2. Summary of Cosmid Clones Isolated

Type	Class I		Class II			
Library	PVG-*RT1*^{a}		PVG-*RT1*^{a}		PVG	
Probes used in screen	D,L		A_β.E_α.E_β		A_α.A_β.$A_{\beta 2}$.E_α.E_β	
Total No. isolated	43		17		27	
No. hybridizing to:	D	27	$A_\alpha + A_\beta$	4	A_α	1
	D + K	4	$A_\beta + A_{\beta 2}$	9	$A_\alpha + A_\beta$	7
	D + L	1	E_α	3	$A_\beta + A_{\beta 2}$	14
	D + L + K	11	E_β	1	E_α	5

of mouse β_2-microglobulin cannot be excluded) and probably represent products of the *RT1.C* region. The reaction with MRC OX18 may be related to cross-reactivity between *RT1.C* and *RT1.A* region products[13] and may be analogous to that seen with MAb between H-2K and a Qa antigen.[14] Restriction digests of the 12 cosmids involved show that they comprise three or four groups with some identical band sizes but with no fragment common to all, implying that several genes are involved. Further restriction mapping will be required to confirm that these cosmids represent cluster(s) of similar genes. Unambiguous mapping of these genes to *RT1.C* should be possible by testing the transformants with specific serologic reagents or by the isolation of low-copy probes from the cosmids and testing on DNA from recombinant rats. The production of new serologic reagents will be necessary to test whether the remaining 31 cosmids cause expression of other, as yet unidentified, rat class I sequences.

Class II Genes

Both PVG-*RT1*^{a} and PVG libraries were screened for class II genes. DNA from clones recovered were tested individually for hybridization to the five probes used. All probes showed clearly positive signals on some clones with negligible hybridization to others. From both libraries, cosmids that hybridized with both A_β and $A_{\beta 2}$ probes were most common, followed by those hybridizing to A_α and A_β. Clones hybridizing to E_α were few, and only one for E_β was recovered from the PVG-*RT1*^{a} library (Table 2). A second screening of the PVG-*RT1*^{a} library using A_α, $A_{\beta 2}$, and E_β probes revealed no additional positive clones. The predominance of clones hybridizing to I-A probes did not result from the presence in rats of multiple I-A–like genes: the restriction maps of the I-A–like clones showed that they were all overlapping and consistent with a single cluster of three genes in the order $A_{\beta 2}$-A_β-A_α. This, together with the recovery of only a single E_β clone from both libraries in three screens with this probe, suggests that the relatively low recovery of I-E–like cosmids is the result of a cloning or stability artifact. Analysis of the E_β cosmid showed two regions that hybridized to the E_β probe with different intensities. The second, weakly hybridizing region was also found in the E_α clones and probably represents the rat analogue of mouse $E_{\beta 2}$.[4] The I-E–like cluster is thus also composed of three genes homologous to E_β-$E_{\beta 2}$-E_α. The spacing of genes in both clusters is very similar to that of the mouse and suggests that the I region of the rat MHC has a virtually identical organization; that is, a total of six genes ordered $A_{\beta 2}$-A_β-A_α-E_β-$E_{\beta 2}$-E_α, spanning a region of about 120 kilobases (kb).[4] The failure to clone DNA segments linking the two clusters (which in the mouse are about 15 kb apart) obviously leaves uncertainties about this conclusion. However, it is consistent with the linkage map for the rat of *RT1.A* – *RT1.B*(=*I-A*) – *RT1.D* (=*I-E*) and with the ordering of the genes for RT1.B β and α chains.[15]

To confirm that the genes isolated were those coding for the serologically identified rat class II antigens, cosmid DNA-transformed L cells were tested with mouse MAb against RT1.B and RT1.D molecules. Intro-

duction of single cosmids carrying both A_α- and A_β-like sequences from either library caused expression of RT1.B as shown by the binding of MAb MRC OX4 and OX6 but not OX17. Cloned cell lines expressing $RT1.B^a$ and $RT1.B^c$ at various levels were successfully obtained. L cells transformed with the E_β + $E_{\beta 2}$-like cloned DNA along with an E_α + $E_{\beta 2}$-like sequence (but not the latter alone) bound MRC OX17 but not OX4 or OX6, indicating that the E_α- and E_β-like genes together encoded $RT1.D^a$. Thus, both *RT1.B* and *RT1.D* regions contain a cluster of three genes: an expressible α and β pair together with a second β-like sequence of unknown function and expressibility.

REFERENCES

1. Howard JC: Metabolism 32:41, 1983 (suppl 1)
2. Gill TJ, Kunz HW, Schaid DJ, et al: J Immunogenet 9:281, 1982
3. Blankenhorn EP, Symington FW, Cramer DV: Immunogenetics 17:475, 1983
4. Hood L, Steinmetz M, Malissen B: Ann Rev Immunol 1:529, 1983
5. Weiss EH, Golden L, Fahrner K, et al: Nature 310:650, 1984
6. Steinmetz M, Winoto A, Minard K, et al: Cell 28:489, 1982
7. Larhammar D, Hammerling U, Denaro M, et al: Cell 34:179, 1983
8. Malissen B, Peele Price M, Goverman JM, et al: Cell 36:319, 1984
9. Diamond AG, Larkins AP, Wright B, et al: Eur J Immunol 14:405, 1984
10. McMaster WR, Williams AF: Immunol Rev 47:117, 1984
11. Fukumoto T, McMaster WR, Williams AF: Eur J Immunol 12:237, 1982
12. Winoto A, Steinmetz M, Hood L: Proc Natl Acad Sci USA 80:3425, 1983
13. Stephenson SP, Livingstone AM, Howard JC, et al: Transplant Proc (this issue)
14. Sharrow SO, Flaherty L, Sachs DH: J Exp Med 159:21, 1984
15. Blankenhorn EP, Cramer DV: Transplant Proc (this issue)

APPENDIX F

THE MURINE T-CELL RECEPTOR USES A LIMITED REPERTOIRE OF EXPRESSED V_β GENE SEGMENTS

Published in

Nature 316, 517-523, 1985

Reprinted from Nature, Vol. 316, No. 6028, pp. 517-523, 8 August 1985

The murine T-cell receptor uses a limited repertoire of expressed V_β gene segments

Richard K. Barth*, Byung S. Kim*†, Nancy C. Lan*, Tim Hunkapiller*, Nancy Sobieck*, Astar Winoto*, Howard Gershenfeld‡, Craig Okada‡, Dan Hansburg§, Irving L. Weissman‡ & Leroy Hood*

* Division of Biology, California Institute of Technology, Pasadena, California 91125, USA
‡ Department of Pathology, Stanford Medical School, Palo Alto, California 94304, USA
§ Department of Pathology, Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111, USA

Only 10 different V_β gene segments were found when the sequences of 15 variable (V_β) genes of the mouse T-cell receptor were examined. From this analysis we calculate that the total number of expressed V_β gene segments may be 21 or fewer, which makes the expressed germline V_β repertoire much smaller than that of the immunoglobulin heavy-chain or light-chain genes. We suggest that β-chain somatic diversification is concentrated at the V_β-D_β-J_β junctions.

BOTH the T-cell receptor and the immunoglobulins are heterogeneous cell-surface glycoproteins that can recognize many antigens[1-3]. It is clear from genomic analysis that they share similar strategies for diversification. T-cell receptor molecules are composed of α- and β-chains, each of which, like the immunoglobulin chains, is divided into variable (V) and constant (C) regions[1-3]. The V regions together form the antigen-binding domain. The β-chain genes of the mouse are the most thoroughly studied T-cell receptor genes. Like the immunoglobulin genes, they are divided into separate V_β, diversity (D_β) and joining (J_β) gene segments that are assembled by recombination during T-cell development to form a V_β gene that is associated with either of two constant ($C_\beta 1$ and $C_\beta 2$) genes[4-8]. There are six functional J_β gene segments clustered just upstream of each C_β gene[4-6] and two D_β gene segments, $D_\beta 1.1$, upstream of the $J_\beta 1$ cluster, and $D_\beta 2.1$, upstream of the $J_\beta 2$ cluster[7,8]. The total number of V_β gene segments is unknown. Like immunoglobulins, the T-cell receptor β-chain has at its disposal three sources of diversity: a multiplicity of germline gene segments[4-9]; combinational diversity through the assembly of different *V*, *D* and *J* segments[4,9-11]; and somatic mutation[9,11]. Immunoglobulin genes have three sources of somatic mutation: junctional flexibility at the sites of gene-segment joining[12-14]; the addition of random nucleotides to either side of the *D*-gene segment in the process of joining (*N*-region diversity)[15]; and somatic hypermutation[16,17]. It is known that β-chain genes use the first two processes but may not use the third[4,7-9,11].

T-cell antigen recognition differs from that mediated by immunoglobulins in that T cells must recognize antigen in the context of a cell-surface molecule encoded by the major histocompatibility complex (MHC), a phenomenon termed MHC restriction[18,19]. T-cytotoxic (T_C) cells, which are capable of killing virus-infected and tumour cells, are mainly restricted by class I gene products of the MHC[20]. T-helper (T_H) cells, which are capable of enhancing B- or T-cell responses, are mainly restricted by class II MHC gene products[21]. T-cell receptor diversification must therefore accommodate antigen recognition and recognition of highly polymorphic determinants on MHC molecules.

To determine the extent of T-cell receptor diversity and its relationship to antigen/MHC recognition, we analysed eight V_β genes from complementary DNA libraries of functional T cells and thymocytes. We have compared these V_β gene sequences with seven from the literature and find that: (1) the expressed V_β gene repertoire is probably small, perhaps less than 21 members; (2) V_β protein segments are structurally similar to immunoglobulin V segments; and (3) there is no simple correlation between antigen and MHC specificity and the use of particular β-chain gene segments.

Expression of V_β gene segments

We determined the nucleotide sequence of eight V_β genes obtained from cDNA libraries that were constructed from thymus cells, the T_H hybridoma 1.9.2 specific for the cytochrome

† Permanent address: Department of Microbiology/Immunology, Northwestern University, Chicago, Illinois 60611, USA.

Table 1 Characteristics and origins of sequenced V_β genes

V_β gene	Class	Strain	Antigen/MHC specificity	V_β	D_β	J_β	Ref.
2B4	T_H	B10.A	Cytochrome *c*/I-$E^k_\alpha E^k_\beta$	3	2.1	2.5	4
1.9.2	T_H	B10.A[5R]	AmDASp/I-E^k (see ref. 22)	1	1.1	1.1	*
3H.25	T_H	C57BL/6	Hen-egg lysozyme/I-A^b	3	1.1	1.2	11
C5	T_H	C57BL/6	Dinitrophenol-ovalbumin/I-A^b	8.1†	2.1	2.5	9
E1	T_H	BALB/c	Trinitrophenol/I-A^d	2	1.1	2.2	9
LB2	T_H	C57BL/6	Chicken red blood cell/I-A^b	6	2.1	2.3	9
HDS11	T_C	BALB.B	H-2^d	7	1.1	2.6	24
AR1	T_C	C57L	H-2^d	2	1.1	2.5	*
86T1	Thymocyte	BALB/c	—	1	1.1	1.3	23
TB2	Thymocyte	C57BL/Ka	—	8.2†	2.1	2.5	*
TB3	Thymocyte	C57BL/Ka	—	4	2.1	2.5	*
TB12	Thymocyte	C57BL/Ka	—	8.1	1.1 or 2.1‡	2.4	*
TB21	Thymocyte	BALB/c	—	5.1	2.1	2.6	*
TB23	Thymocyte	BALB/c	—	8.3†	ND	ND	*
BW5147	Tumour	AKR	—	1	2.1	2.5	*

* This paper.
† The three members of the $V_\beta 8$-subfamily are denoted $V_\beta 8.1$, $V_\beta 8.2$ and $V_\beta 8.3$.
‡ So little of the D_β gene segment remains in the rearranged V_β gene that it is impossible to know which D_β gene segment contributed the sequence. ND, not determined.

2

c peptide AmDASp (ref. 22) and a functional T_C-cell line specific for MHC alloantigens. The sequence of these V_β genes and seven additional V_β genes taken from the literature were then compared (Fig. 1*a*, Table 1)[4,9,11,23,24]. Of the 15 V_β genes analysed, 6 are from T_H cells, 2 from T_C cells, 6 from thymocytes and 1 from a T-cell tumour (Table 1). The antigen/MHC specificities of the functional cells are listed in Table 1. The specific functions and antigen/MHC specificities of the T-cell tumour and the cells that generated the thymocyte-derived V_β genes are unknown. As 99% of thymocytes do not migrate out of the thymus[25], some of the thymocyte-derived V_β genes may come from non-functional T cells. The 10 distinct protein sequences of the 15 V_β gene segments are shown in Fig. 1*b*: there are 10 distinct sequences. For reasons discussed below, the first seven sequences listed in Fig. 1*b* are designated V_β1–7 and the last three V_β8.1, V_β8.2 and V_β8.3. One V_β gene segment appears three times (V_β1), three appear twice (V_β2, V_β3, and V_β8.1) and six appear once (V_β4, V_β5.1, V_β6, V_β7, V_β8.2 and V_β8.3). Moreover, when the sequences of seven additional V_β genes isolated from six T_H cells by our laboratory[26] and others are included (J. Goverman, unpublished data; M. Steinmetz, personal communication; A. Fotedar, P. Morinaga, B. Singh, T. Tamaoki and T. Wegmann, personal communication; S. Hedrick, personal communication), 11 different V_β gene segments are found in 22 rearranged genes. Six of the eleven different V_β gene segments appear more than once. Common V_β gene segments occur in functional T cells, thymocytes and the T-cell tumour BW5147. The repeated use of V_β gene segments in a sample of this size suggests that the repertoire of expressed V_β gene segments is small. If we assume that the V_β gene segments are expressed randomly, the size of the V_β gene segment repertoire can be determined statistically.

If there are L species of V_β gene segments and they are expressed with equal probability and are randomly selected

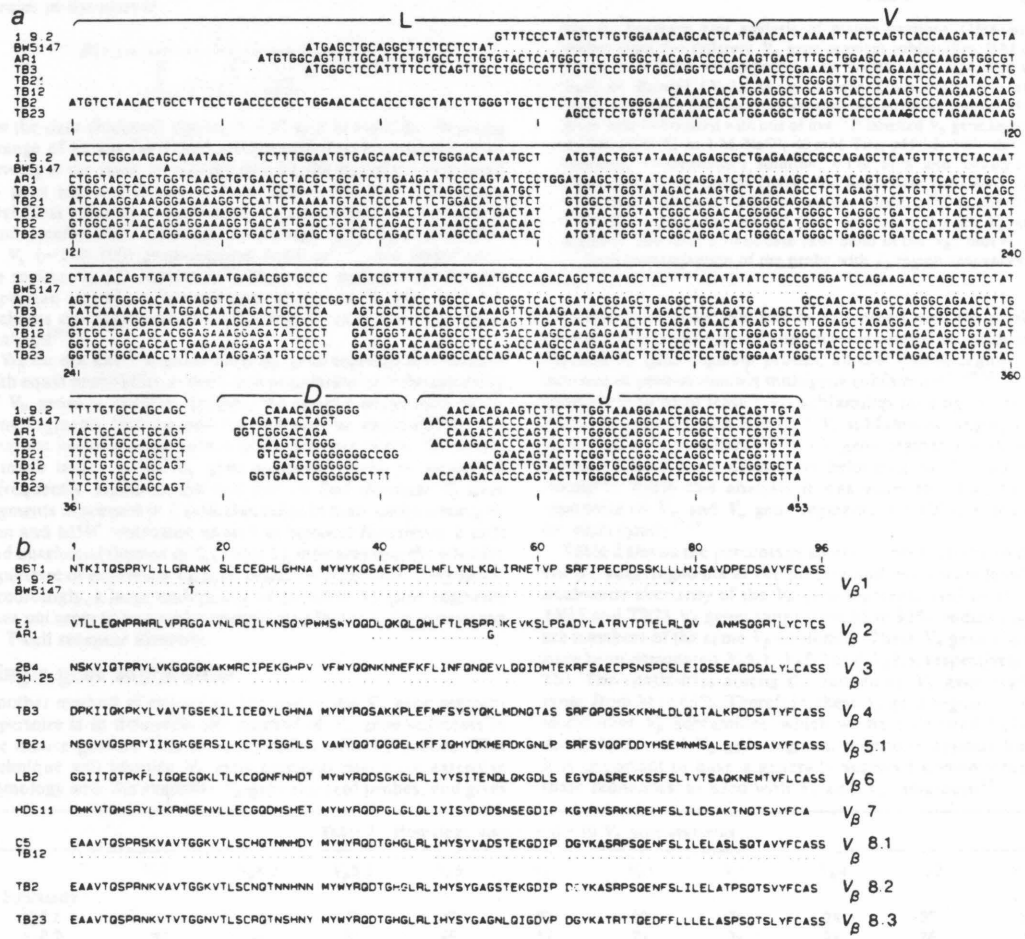


Fig. 1 *a*, DNA sequence of eight V_β genes. All cDNA libraries were constructed in the λgt10 cloning vector by a modification of the method of Huynh *et al.*[56]. Libraries were screened with a ^{32}P-labelled C_β cDNA probe, positive clones were isolated and cDNA inserts were subcloned into the M13 mp8 or mp10 vectors. The V_β genes of each subclone were sequenced on both strands by a modification of the dideoxynucleotide-chain terminating method of Sanger[57], using specific oligonucleotide primers[58]. The sequences are displayed with the V_β and J_β gene segments and *D* regions aligned separately. L, nucleotides encoding the leader peptide. The TB23 subclones contained only a V_β gene segment. Dots, identity between the V_β segments used by 1.9.2 and BW5147. The cDNA obtained from AR1 contained only a portion of a V_β gene segment beginning at position 161. For purposes of sequence alignment, the 5′ portion of the homologous V_β gene segment published previously[9] has been added. *b*, Translated protein sequences of 15 V_β gene segments. The DNA sequences of the eight V_β gene segments shown in *a*, as well as seven V_β gene segments published previously, were translated and aligned to each other to maximize sequence homology. (See Table 1 for the source of each sequence.)

from the pool, then the probability (P) of obtaining an observed result is

$$P(m_1, m_2, \ldots, m_N) = \frac{1}{L^N} \frac{L!}{(L-M)!M!} \times \frac{N!}{(1!)^{m_1}(2!)^{m_2}\ldots(N!)^{m_N}} \times \frac{M!}{m_1!m_2!\cdots m_N!}$$

where m_1 is the number of different species found once; m_2, the number of different species found twice; m_N, the number of different species found N times; $N = \sum_{i=1}^{N} im$ = the total number of samples analysed; and $M = \sum_{i=1}^{N} m_i$ = the number of different species found in the sampling. Conversely, given the observed result, the relative likelihood that there are exactly L species in the pool is

$$P(L) = \frac{1}{\sum_{L'=M}^{\infty} \frac{L'!}{L'^N(L'-M)!}} \frac{L!}{L^N(L-M)!}$$

For the data discussed above, $N = 22$ and $M = 11$. By choosing a range of values for L and testing each for the probability of arriving at the value $M = 11$ for $N = 22$, the size of the expressed V_β gene segment family is 21 or less at the 95% confidence level. This estimate of the mouse V_β-gene segment repertoire is much smaller than the mouse immunoglobulin V_H (~100-300) or V_κ (~100-300) gene-segment families[27,28], but larger than the mouse V_λ (2) repertoire[29]. However, mouse λ-chains are expressed in only a few per cent of mature B cells, whereas β-chains are expressed in all T_H and T_C cells that have been analysed[30,31].

We do not know whether each V_β gene segment is expressed with equal probability in the T-cell population or if the sampling of V_β genes is random. In fact, the relative occurrence of V_β gene segments that we observe can also be explained by the frequent use of a small subset of V_β gene segments. Therefore, a much larger set of V_β gene segments could be expressed infrequently. However, the fact that we find identical V_β gene segments expressed in T cells that differ in their antigen recognition and MHC restriction as well as between functional T cells and unselected thymus cells (Table 1), indicates that the effective repertoire of expressed V_β gene segments is probably very small. Accordingly, a large multiplicity of germline V_β gene segments does not seem to be a major contributing factor in the generation of T-cell receptor diversity.

Single gene subfamilies

Another method of estimating the size of the V_β gene segment repertoire is to determine the number of V_β gene segments in the mouse genome that cross-hybridize with V_β probes. This technique will identify V_β gene segments that have extensive homology with the available V_β gene segment probes, and gives a minimum estimate that can be compared with the predicted number. When this type of analysis was carried out with mouse V_H and V_κ gene segment probes, all the V gene segments fell into one of several distinct multigene subfamilies[27,28,32]. In mice there seem to be at least 7 V_H subfamilies ranging in size from 2 to >40 members and at least 5 V_κ subfamilies ranging in size from 2 to >20 members[27,28,32]. (V gene segments with ≥75% similarity have been defined as belonging to the same subfamily[33].) From this analysis it was estimated that the total repertoire of V_H and V_κ gene segments is ~100-300 members for each family[27,28].

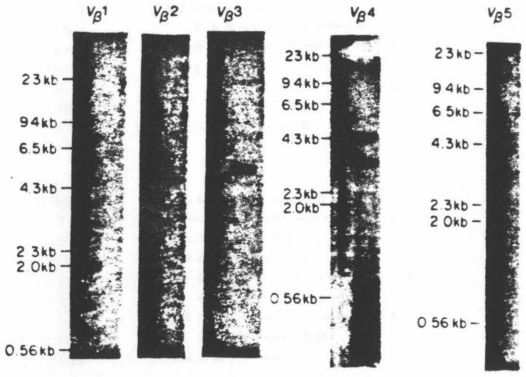


Fig. 2 Southern blot analysis of mouse germline DNA using probes from five different V_β gene segment subfamilies. BALB/c mouse liver DNA was digested with *Eco*RI ($V_\beta1$, $V_\beta4$, $V_\beta5$ and $V_\beta8$) or *Bam*HI ($V_\beta2$). 10 μg of each DNA digest was electrophoresed on a 0.8% agarose gel and transferred to nitrocellulose. Blots were hybridized with one of five ^{32}P-labelled V_β gene segment probes at 65 °C in 1 M NaCl, 40 mM Tris, pH 7.5, 10% dextran sulphate, 1 × Denhardt's solution, 0.1% SDS and 100 μg ml^{-1} denatured salmon-sperm DNA. The blots were then washed at 65 °C in 0.15 M NaCl, 0.015 M sodium citrate, pH 7.5, and 0.1% SDS. The probes are derived from subclones of each V_β gene segment. The faint 5.7-kilobase (kb) band in the $V_\beta2$ blot results from contamination of the probe with J_β region sequences.

Table 2 shows the percentage similarity between the 10 different V_β gene segments at the protein and nucleotide level. The nucleotide similarity of the V_β gene segments used by the TB2, TB12 and TB23 V_β genes range from 85 to 92%, indicating they are members of the same V_β subfamily. These V_β gene segments have been designated $V_\beta8.1$, $V_\beta8.2$ and $V_\beta8.3$, respectively (Fig. 1*b*). The similarities among the remaining V_β gene segments range from 34 to 63%. Therefore, these V_β gene segments belong to different V_β subfamilies, which we have denoted $V_\beta1$-$V_\beta7$. As the expressed V_β gene segment repertoire appears limited, it is important to have a generally accepted nomenclature for these sequences, as used with V_L and V_H subgroups[34].

Table 2 Homology matrix of the 10 V_β gene segments

	$V_\beta8.1$	$V_\beta8.2$	$V_\beta8.3$	$V_\beta6$	$V_\beta7$	$V_\beta1$	$V_\beta3$	$V_\beta4$	$V_\beta2$	$V_\beta5$
Subfamily										
$V_\beta8.1$	—	90	77	47	52	29	26	28	27	29
$V_\beta8.2$	92	—	81	45	52	29	24	28	26	26
$V_\beta8.3$	85	88	—	41	48	28	23	28	29	24
$V_\beta6$	55	54	54	—	43	27	27	29	25	22
$V_\beta7$	61	63	60	53	—	28	23	27	23	29
$V_\beta1$	45	43	44	45	49	—	33	53	20	37
$V_\beta3$	46	47	47	46	48	50	—	30	18	37
$V_\beta4$	46	46	46	47	47	62	50	—	20	33
$V_\beta2$	42	44	42	39	39	34	38	38	—	18
$V_\beta5$	45	46	43	41	48	53	54	47	38	—

Numbers above the diagonal designate the percentage similarity of sequences on the x and y axes when compared at the protein level; numbers below the diagonal show percentage similarity at the DNA level.

To determine the size of the different V_β subfamilies, Southern blot analysis was performed on mouse liver DNA using DNA probes for the $V_\beta 1$, $V_\beta 2$, $V_\beta 4$, $V_\beta 5$ and $V_\beta 8$ subfamilies (Fig. 2). Three of the V_β probes show a single band, indicating that each V_β gene segment represents a different single-gene segment subfamily ($V_\beta 1$, $V_\beta 2$ and $V_\beta 4$). The $V_\beta 5$ and $V_\beta 8$ subfamilies appear to have two and three members, respectively (Fig. 2). Because the V_β gene segment used by the TB21 V_β gene is the first isolated member of the $V_\beta 5$ subfamily, we have denoted it $V_\beta 5.1$ (Fig. 1*b*). It has been previously reported that the $V_\beta 1$, $V_\beta 2$, $V_\beta 3$, $V_\beta 6$ and $V_\beta 7$ subfamilies are single genes and that the $V_\beta 8$ subfamily has two members[9,11,24]. Thus, six different V_β subfamilies with one member, one V_β gene segment subfamily with two members and one V_β subfamily with three members have been identified. Including the additional V_β gene segment characterized by Malissen *et al.*[26], there are at least 12 mouse V_β gene segments. This minimum estimate for the size of the V_β gene segment family falls within the range indicated by the statistical analysis presented above and is consistent with the hypothesis that the V_β gene segment repertoire is small.

The V_β gene segment family with its six single-member subfamilies differs from those of the immunoglobulin V_H and V_κ gene families, each of which contains 5 or more subfamilies with 2–40 or more members[27,28,32]. Southern blots of DNAs from several rodent species as well as rabbit and human analysed with various V_β probes indicate that the single-member V_β subfamily sequences are less conserved between species than those of the three-member family ($V_\beta 8$)[9]. It has been suggested that this difference reflects selective pressures on the singie-copy sequences to diverge rapidly, presumably to accommodate recognition of antigen in a changing MHC context[9]. An *ad hoc* argument for specific, positive selection must then be made to explain the unique conservation of the $V_\beta 8$ subfamily. In apparent conflict with this view is the observation that very little restriction enzyme polymorphism of single copy V_β sequences is seen between mouse strains that have diverged significantly in their MHC genes (B.S.K. and R.K.B., unpublished observations; D. Loh, personal communication). In addition, the inbred SJL mouse has deleted two of the five V_β gene segment subfamilies examined (B.S.K. and R.K.B., unpublished observations; D. Loh, personal communication). This indicates that mice have distinct V_β haplotypes containing different combinations of V_β subfamilies. Hence, an alternative explanation for the lack of V_β-gene segment conservation between species is that the ancestor to mammals contained different V_β haplotypes and during speciation distinct V_β haplotypes were passed on to different evolutionary lines. This model does not require a high rate of V_β gene segment mutation and consistent with the lack of restriction enzyme V_β polymorphism in mice. Additional data on the evolutionary divergence of the V_β gene segments should clarify the explanations for these observations.

It is unclear why single-member subfamilies seem to have arisen exclusively in the V_β gene family. Note that no V_β pseudogenes have been found, whereas at least 30% of V_H gene segments are pseudogenes[27]. These observations raise a question about the mechanisms that are responsible for retarding the duplicative processes seen in other *V* gene families.

Diversification mechanisms

Despite the fact that a limited number of V_β gene segments have been identified, all 15 of the V_β gene sequences examined are distinct from one another because of combinatorial and somatic mutational processes (Fig. 1*a*)[4,7–11]. The germline, combinatorial and somatic mutation contributions to V_β gene assembly are summarized below.

Germline. The β-chain gene family differs from its V_H counterpart in the apparently limited number of germline V_β and D_β gene segments, the expressed V_β gene segment repertoire consisting of perhaps 21 or fewer members compared with 100–300 germline V_H gene segments. In addition, although we do not know the number of germline D_β gene segments, comparisons

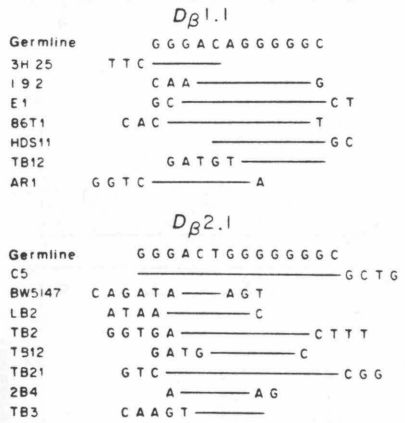


Fig. 3 The *D* regions from 15 V_β genes may arise from just two D_β gene segments. The *D* regions shown in Fig. 1*a* and those published previously[4,9,11,23,24] were aligned to either $D_\beta 1.1$ or $D_\beta 2.1$. The regions homologous to either D_β gene segment are represented by a straight line. The additional nucleotides flanking the germline D_β sequences are presumed to be added by *N*-region diversity. The *D* region used by TB12 can be derived from either $D_\beta 1.1$ or $D_\beta 2.1$ and is so indicated.

of the *D* segments found in rearranged V_β genes indicates that all could be derived from the two D_β gene segments previously identified[7,8] assuming extensive junctional flexibility and *N*-region diversification (Fig. 3). The heavy-chain locus, by contrast, has at least 10–20 D_H segments[35]. Finally, there are 12 apparently functional J_β gene segments, 6 in each J_β gene cluster[4–6], compared with 4 J_H gene segments[36–38].

Combinatorial. Combinatorial joining permits either D_β gene segment to be joined to any downstream J_β gene segment ($6\ D_\beta 1 J_\beta 1 + 6 D_\beta 1 J_\beta 2 + 6 D_\beta 2 J_\beta 2 = 18 D_\beta\text{-}J_\beta$ rearrangements). Individual V_β gene segments appear to join any D_β-J_β rearrangements ($21 \times 18 = 378$ V_β genes). The $D_\beta 1$ and $D_\beta 2$ sequences are each used approximately half the time in the sample analysed and $D_\beta 1$-$J_\beta 2$ joinings occur as frequently as $D_\beta 1$-$J_\beta 1$ joinings (3 against 3)(Table 1). Eight different $J_\beta 1$ and $J_\beta 2$ gene segments are used, with $J_\beta 2$ gene segments being used in 11 out of 14 examples. Thus, one would expect individual V_β gene segments to join with the $J_\beta 1$ gene cluster 25% of the time and the $J_\beta 2$ gene cluster 75% of the time, which is what is found, supporting the contention that the joining of the $D_\beta 1.1$ gene segment to either J_β cluster occurs randomly. However, the $J_\beta 2.5$ gene segment is used 6 of the 11 times that the $J_\beta 2$ cluster is used. Whether this bias represents the selective effects of antigen or the mechanisms involved in DNA rearrangement is uncertain.

There may be two additional combinatorial mechanisms. Asymmetrical recognition sequences surrounding the D_β gene segments potentially permit V_β-J_β and D_β-D_β joinings[7,8]. Data consistent with the former possibility have been presented[39], although the interpretation of these data is difficult because of the possible loss of *D* gene sequences during V_β gene segment rearrangement. As yet there is no evidence for either mechanism; hence, if these joinings do occur, they are infrequent.

Somatic mutation. Junctional flexibility in joining gene segments is illustrated in Figs 1*a* and 3. One to six extra nucleotides are found at either end of the *D* gene segments. Interestingly, there appears to be no G/C bias (50%) in this *N*-region diversification in contrast to that reported for immunoglobulin *N*-region diversity[15].

The D_β gene segments may join to the V_β gene segments with equal probability in all three translational reading frames, with

the only requirement being that the D_β-J_β and V_β-D_β joinings leave the J_β sequence in the proper translational frame (data not shown)[11]. In contrast, examination of >75 productively rearranged V_H genes indicate that D_H-gene segment families show a strong preference for joining in one translational reading frame (85%)(M. Kaartinen and O. Mäkelä, personal communication; T.H., unpublished observations). Presumably this difference arises as a consequence of more rigid structural or selective constraints on the D_H as opposed to D_β segments.

The somatic hypermutation of immunoglobulin genes occurs late in B-cell development, perhaps on exposure to antigen[40,41]. By contrast, two V_β gene segments expressed in functional T_H cells specific for different antigens are identical to the germline sequence[4,11]. On the other hand somatic variants can arise in alloreactive T cells *in vitro*[42], although the physiological relevance of this observation is uncertain. We see two nucleotide substitutions and one replacement among the three V_β1, and two nucleotide substitutions and one replacement between the two V_β2 gene segments that have been sequenced (Fig. 1*a*, *b*)[9,23]. As all these V_β sequences are derived from different strains of mice (Table 1), these differences may result from polymorphism. No other differences were found and we tentatively conclude, in agreement with earlier workers[4,11], that somatic hypermutation, if it exists at all, is much less extensive in V_β than in immunoglobulin *V* genes.

V-region similarities

It has been suggested that both the total diversity and the distribution of variability of V_β segments differ from those of immunoglobulin V segments and that this may reflect the additional requirements for MHC recognition[9]. We have analysed the pattern of sequence diversity of the available V_β sequences with this in mind.

The percentage similarities at the protein level (Table 2) between the different V_β subfamilies ranges from 18 to 53% and 77 to 90% between members of the V_β8 subfamily. It has been suggested that V_β segments are substantially more divergent than V_H sequences[9]. Although their range extends to a lower value (18%) than that observed so far for known mouse V_H segments (34%), or human V_H segments (24%)(data not shown), when sampling biases are considered (see below) the maximum variation between different V_β segments (82%) does not seem significantly larger than the maximum variation found between immunoglobulin V subfamilies (76%).

In immunoglobulin V regions, most of the sequence variability is clustered within three specific 'hypervariable' regions that form the antigen-binding crevice of antibody molecules[43–46]. Two of these hypervariable regions are encoded by the *V* gene segments and the third is found in the V_H-D_H-J_H or V_L-J_L junction regions[47]. We have compared the pattern of variability of the 10 translated mouse V_β gene segments (Fig. 1*b*) with that of a set representing all the known human V_H segments and a set of 18 human V_H segments with blocked α-amino groups (Fig. 4). Human V_H segments were chosen for comparison because they offer a more random representation of V_H sequences than any set of mouse *V* segments that has been sequenced. This is because mouse V_H sequences have been highly selected in comparison with V_β sequences in two ways. First, mouse V_H sequences are derived mainly from immunoglobulins that recognize a relatively limited number of antigens. Second, for technical reasons, mouse V_H sequences are almost exclusively determined from heavy chains with unblocked α-amino groups despite the fact that 80% of mouse serum immunoglobulins have heavy chains with blocked α-amino groups[48]. The blocked human V_H sequences, on the other hand, were selected randomly from various tumours and patients with other pathological conditions[47]. We have found that the variability distribution of the mouse V_β segments (Fig. 4*a*) is very similar to the distribution found for the set of human α-amino blocked V_H segments (Fig. 4*b*). The variability distribution of the total set of human V_H segments (Fig. 4*c*) represents a less random sampling than the human α-amino-blocked V_H segments for the same reasons

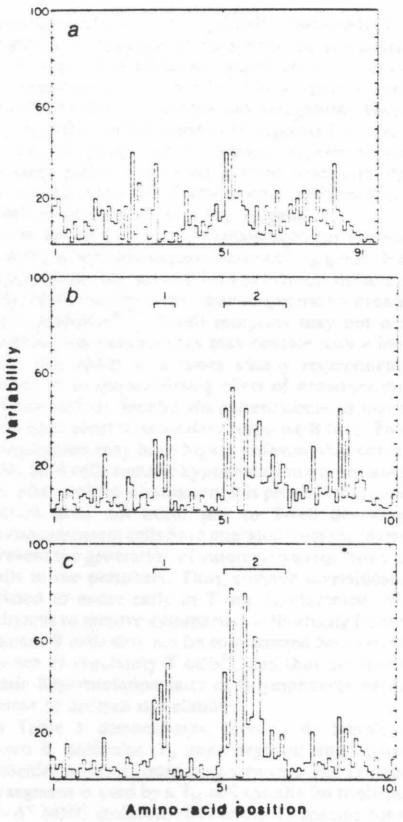


Fig. 4 Variability plot of V_β and V_H segments. Variability at each amino-acid position *N* is calculated[43] as: variability$_N$ = no. of different amino acids that occur at *N*/frequency of most commonly occurring amino acid at *N*. Hypervariable regions can be defined empirically as a set of residue positions whose average variability is substantially greater than the mean variability of the entire sequence. The sequences used are all those available from the protein information resource of the National Biomedical Resource Foundation and GenBank. *a*, Translated sequence of the 10 distinct V_β gene segments shown in Fig. 1*b*; *b*, 18 α-amino-blocked human V_H sequences; *c*, 31 human V_H sequences representing sequences blocked and unblocked at the α-amino position.

given for the mouse sequences, and shows a more accentuated variability at the two classically defined hypervariable regions. The total set of human V_H sequences also exhibits a slightly lower background variability, partly resulting from the larger size of the sampling. Thus, it is critical to consider sample size and selection when this type of analysis is conducted. We conclude that the distributions of variability in the V_β and V_H segments are not significantly different from one another. Our results are not consistent with the suggestion[9] that V_β regions have novel hypervariable regions relative to immunoglobulin V regions.

We have also compared V_β and immunoglobulin V segments by analysing them for two properties believed to reflect important structural features of these molecules, the distribution of β-pleated sheet-forming potential[49] and the predicted hydrophobicity profile[50]. We find the results of these analyses to be almost identical for mouse V_β, V_H and V_κ segments (Fig. 5). Patten *et al.* have also reported that the β-pleated sheet patterns of several V_β segments conform to that of representative V_H and V_κ sequences[9]. V_β sequences also conserve essentially the

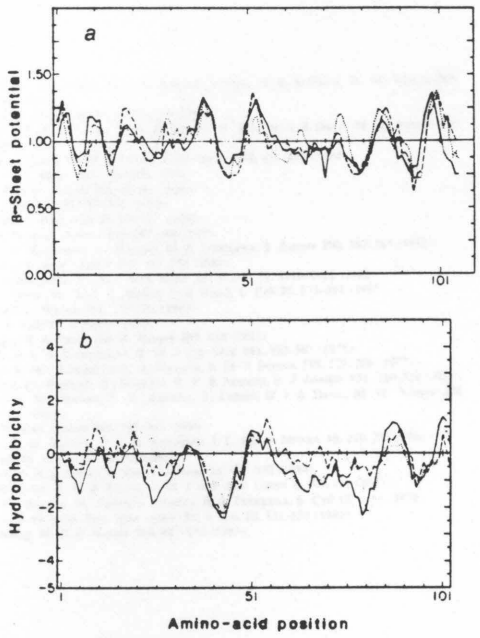


Fig. 5 Secondary structure analyses of V_β, V_H and V_κ segments. *a*, β-pleated sheet potential plots, using the method of Chou and Fassman[49]; *b*, hydrophobicity plots using the scale of Kyte and Doolittle[50]. Solid lines, β; dotted lines, H; dashed lines, κ. Analyses in *a* and *b* were based on the average value at each position for 55 V_H regions, 100 V_κ regions and 10 V_β regions.

same group of residues that in immunoglobulins are thought to be important in intra-chain structural interactions (data not shown)[51,52]. These results strongly support the contention that the general biochemical characteristics and predicted secondary structures of the V_β, V_H and V_κ regions are very similar to one another. Accordingly, we predict that the T-cell receptor and immunoglobulin molecules fold into comparable tertiary structures. Therefore, from the analysis of the sequence variability and structural predictions, we find no evidence for the existence of any fundamental differences in how V_β gene segments can contribute to determinant recognition when compared with immunoglobulin *V* gene segments. This conclusion is supported by the observation that MHC-restricted antibodies have been raised to influenza antigens[53]. This observation implies that there need be nothing structurally unique about the T-cell receptor structure and its ability to recognize antigen and MHC-restricting elements.

Antigen/MHC specificity

T cells can recognize a range of antigens similar to those recognized by B cells, but in conjunction with the entire range of polymorphic MHC molecules found in a species. Therefore, the T-cell receptor repertoire is expected to be at least equal to the immunoglobulin repertoire. Our observations suggest that the β-chain genes use fewer *V* gene segments than do immunoglobulins and that somatic hypermutation is infrequent or even non-existent. On the other hand, the V_β subfamilies exhibit the same overall range of sequence diversity as different V_H subfamilies. Furthermore, the contribution of J_β gene segments is three times that of J_H or J_κ gene segments, and because the D_β gene segments can be used readily in all three translational reading frames, the smaller family size than is seen in immunoglobulins is at least partially compensated for. Finally, β-chains and immunoglobulins both use junctional flexibility and *N*-region diversification, which are major contributors to the generation of diversity within a specific region of the molecule involved in determinant recognition. Hence, we conclude that the limited number of expressed V_β gene segments and the infrequent use of somatic hypermutation does not necessarily reflect a restricted β-chain repertoire. Rather, these observations imply that β-chain somatic diversification is more focused to the 3' portion of the V_β gene.

There are three possible explanations for the apparent low frequency of somatic hypermutation of V_β genes. First, somatic hypermutation may act in B cells to increase the antigen binding affinity of antibodies rather than to generate a broader range of antigen response[41,54]. T-cell receptors may not need somatic hypermutation because they may operate with a lower binding affinity due either to a lower affinity requirement for T-cell response or to the stabilizing effect of accessory molecules on the T-cell surface. Second, the current sampling may not include T cells equivalent to secondary-response B cells. Third, somatic hypermutation may have highly unfavourable consequences in T cells. In B cells somatic hypermutation occurs late in development, after antigen stimulation. It is possible that somatic hypermutation does not occur late in T-cell development, after immunocompetent cells have migrated from the thymus, in order to prevent the generation of autoreactive regulatory or cytotoxic T cells in the periphery. Thus, somatic diversification may be restricted to occur early in T-cell development, which allows the thymus to remove autoreactive cells arising from the somatic variation. B cells may not be so restricted because of the strong influence of regulatory T cells[55], and thus are free to undergo somatic hypermutation later in B-lymphocyte development in response to antigen stimulation.

As Table 1 demonstrates, there is no simple correlation between a particular V_β gene segment and distinct antigen specificities or MHC-restricting elements. For example, the $V_\beta 2$ gene segment is used by a T_H-cell specific for trinitrophenol and the I-A^d MHC molecule (E1) and a T_C specific for the H-$2D^d$ alloantigen (AR1). If the V_β and V_α regions fold in a manner similar to that of their immunoglobulin counterparts, as is suggested for V_β regions by our earlier analysis, then both chains will play a critical role in generating the binding site for antigen plus MHC. Accordingly, there is no reason to believe that either chain will have a particular role in recognizing either antigen or MHC individually.

We thank Marcus Meister for deriving the statistical calculation used for V_β gene segment number estimates and help with statistical analyses. We also thank Susan Stone, Debbie Maloney and Marty Garcia for technical assistance, Gerald Siu and Dr Mitchell Kronenberg for helpful advice and stimulating discussion, Drs Nilabh Shastri, Joan Kobori, Nancy Costlow, Joan Goverman, Stephen Hunt, Regina Haars, Ulf Landegren and Martha Zuniga for critical comments, and Gwen Anastasi, Connie Katz and Bertha Jones for help in preparing the manuscript. We acknowledge the following sources of support: the Cancer Research Institute (R.K.B.) and the NIH (L.E.H.).

Note added in proof: Recently, the sequences of 15 additional V_β genes have been determined (D. Loh, personal communication). When these are added to the 22 existing V_β sequences, only 16 different V_β gene segments are defined from a total of 37 sequences. These additional data support our hypothesis of a limited V_β gene segment repertoire. Furthermore, when these data are included with our own data and analysed for the distribution of variability, hydrophobicity and β-pleated sheet-forming potential, the results are very similar to those presented here.

Received 10 April; accepted 13 June 1985.

1. Allison, J. P., McIntyre, B. & Bloch, D. *J. Immun.* **129**, 2293-2300 (1982).
2. Haskins, K. *et al. J. exp. Med.* **157**, 1149-1169 (1983).
3. Meuer, S. C. *et al. J. exp. Med.* **157**, 705-719 (1983).

4 Chien, Y.-H., Gascoigne, N. R. J., Kavaler, J., Lee, N. E. & Davis, M. M. *Nature* **309**, 322-326 (1984).
5 Malissen, M. *et al. Cell* **37**, 1101-1110 (1984).
6 Gascoigne, N. R. J., Chien, Y.-H., Becker, D. M., Kavaler, J. & Davis, M. M. *Nature* **310**, 387-391 (1984).
7 Kavaler, J., Davis, M. M. & Chien, Y.-H. *nature* **310**, 421-423 (1984).
8 Siu, G. *et al. Nature* **311**, 344-349 (1984).
9 Patten, P. *et al. Nature* **312**, 40-46 (1984).
10 Siu, G. *et al. Cell* **37**, 393-401 (1984).
11 Goverman, J. *et al. Cell* **40**, 859-867 (1985).
12 Weigert, M. *et al. Nature* **283**, 497-499 (1980).
13 Sakano, H., Kurosawa, Y., Weigert, M. & Tonegawa, S. *Nature* **290**, 562-565 (1981).
14 Kurosawa, Y. *et al. Nature* **290**, 565-570 (1981).
15 Alt, F. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4118-4122 (1982).
16 Kim, S., Davis, M., Sinn, E., Patten, P. & Hood, L. *Cell* **27**, 573-581 (1981).
17 Tonegawa, S. *Nature* **302**, 575-581 (1983).
18 Golub, E. *Cell* **21**, 603-604 (1980).
19 Matzinger, P. & Zamoyska, R. *Nature* **297**, 628 (1982).
20 Doherty, P. C. & Zinkernagel, R. M. *J. exp. Med.* **141**, 502-507 (1975).
21 Thomas, D. W., Yamashita, U. & Shevach, E. M. *J. Immun.* **119**, 223-226 (1977).
22 Hansburg, D., Fairwell, T., Schwarz, R. H. & Appella, E. *J. Immun.* **131**, 319-324 (1983).
23 Hedrick, S. M., Nielsen, E. A., Kavaler, J., Cohen, D. I. & Davis, M. M. *Nature* **308**, 153-158 (1984).
24 Saito, H. *et al. Nature* **309**, 757-762 (1984).
25 Scollay, R. G., Butcher, E. C. & Weissman, I. L. *Eur. J. Immun.* **10**, 210-218 (1980).
26 Malissen, M. *et al. Cell* (in the press).
27 Brodeur, P. H. & Riblet, R. *Eur. J. Immun.* **14**, 922-930 (1984).
28 Cory, S., Tyler, B. M. & Adams, J. M. *J. mol. appl. Genet.* **1**, 103-116 (1981).
29 Brack, C., Hirama, M., Lenhard-Schuller, R. & Tonegawa, S. *Cell* **15**, 1-14 (1978).
30 Hedrick, S. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 531-535 (1985).
31 Kronenberg, M. *et al. Nature* **313**, 647-653 (1985).
32 Dildrop, R. *Immun. Today* **5**, 85-86 (1984).
33 Crews, S., Griffin, J., Huang, H., Calame, K. & Hood, L. *Cell* **25**, 59-66 (1981).
34 Potter, M. *Adv. Immun.* **25**, 141-211 (1977).
35 Kurosawa, Y. & Tonegawa, S. *J. exp. Med.* **155**, 201-218 (1982).
36 Early, P., Huang, H., Davis, M., Calame, K. & Hood, L. *Cell* **19**, 981-992 (1980).
37 Bernard, O. & Gough, N. M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3630-3634 (1980).
38 Gough, N. M. & Bernard, O. *Proc. natn. Acad. Sci. U.S.A.* **78**, 509-513 (1981).
39 Yoshikai, Y. *et al. Nature* **312**, 521-524 (1984).
40 Manser, T., Huang, S.-Y. & Gefter, M. L. *Science* **226**, 1283-1288 (1984).
41 Griffith, G. M., Berek, C., Kaartinen, M. & Milstein, C. *Nature* **312**, 271-275 (1984).
42 Augustin, A. A. & Sim, G. K. *Cell* **39**, 5-12 (1984).
43 Wu, T. T. & Kabat, E. A. *J. exp. Med.* **132**, 211-250 (1970).
44 Amzel, L. M. & Poljak, R. J. *Ann. Rev. Biochem.* **48**, 961-997 (1979).
45 Davies, D. R. & Metzger, H. *Ann. Rev. Immun.* **1**, 87-117 (1983).
46 Amit, A. G., Mariuzza, R. A., Phillips, S. E. V. & Poljak, R. J. *Nature* **313**, 156-158 (1985).
47 Kabat, E. A., Wu, T. T., Bilofsky, H., Reid-Miller, M. & Perry, H. *Sequences of Immunological Interest* (U.S. Department of Health and Human Services, Washington, DC, 1983).
48 Capra, J. D., Wasserman, R. L. & Kehoe, J. M. *J. exp. Med.* **138**, 410-427 (1973).
49 Chou, P. & Fassman, G. *Ann. Rev. Biochem.* **47**, 251-276 (1978).
50 Kyte, J. & Doolittle, R. F. *J. molec. Biol.* **157**, 105-132 (1982).
51 Poljak, R. J., Amzel, L. M., Chen, B. L., Phizackerley, R. P. & Saul, F. *Proc. natn. Acad. Sci. U.S.A.* **71**, 3440-3444 (1974).
52 Saul, F. A., Amzel, L. M. & Poljak, R. J. *J. biol. Chem.* **253**, 585-597 (1978).
53 Wylie, D. E., Sherman, L. A. & Klinman, N. R. *J. exp. Med.* **155**, 403-414 (1982).
54 Rodwell, J., Gearhart, P. & Karush, F. *J. Immun.* **130**, 313-316 (1983).
55 Mitchell, G. F. & Miller, J. F. A. P. *J. exp. Med.* **128**, 821-837 (1968).
56 Huynh, T. V., Young, R. A. & Davis, R. W. in *DNA Cloning: A Practical Approach* (ed. Glover, D.) (IRL, Oxford, 1984).
57 Sanger, F., Nicklen, S. & Coulsen, A. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
58 Strauss, E. C., Kobori, J. A., Siu, G. & Hood, L. E. *Analyt. Biochem.* (in the press).