MOLECULAR CLONING OF HUMAN HLA-DS AND MURINE THY-1 ANTIGENS: MEMBRANE PROTEINS WITH HOMOLOGY TO IMMUNOGLOBULINS

BY

Hsiu-Ching Chang

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirments
for the degree of

DOCTOR OF PHILOSOPHY

Department of Microbiology & Public Health

1984

TO MY HUSBAND, YEN-MING HSU AND OUR PARENTS

ACKNOWLEDGEMENTS

I would like to express my deeply appreciation to Dr. Jack Silver for his advice, encouragement and most imortant, the financial support during my entire graduate study. I am also grateful to Drs. Jerry Dodgson, Harold L. Sadoff, John L. Wang, Walter Esselman and Susan Conrad for serving as my research Guidance Committee. Specially, I thank Jerry for teaching me the molecular cloning technology, and Susan for giving me the space to finish up my thesis and being my advisor after Jack left M.S.U.

I would like to express my gratitude to Dr. Tetsuya Moriuchi for his collaborative efforts that aided in the progress of this work, and also I would like to thank Drs. Tetsunori Seki and Roger Denome for their help and allowing me to use their unpublished data here.

TABLE OF CONTENTS

Page

LIST OF TABLES vi

LIST OF FIGURES vii

ABBREVIATIONS ix

INTRODUCTION 1

CHAPTER I LITERATURE REVIEW 3

I. The major histocompatibility complex in mouse and man 3

General introduction of the MHC 3
The H-2 complex and the human leuckocyte antigens (HLA) 4
Structure and function of the MHC antigens 6

II. Class II antigens of the MHC 8

The molecular structure of class II antigens 8
The isolation of genes coding for class II antigens 9
Polypeptides of class II antigens 9
Exon and intron organization of class II genes 14
Genetic map of the class II region of the MHC 16
Polymorphism of class II antigens 17
Biosynthesis of class II antigens 19
Gene expression and functional studies of class II genes . . . 20

III. The Thy-1 antigen--a membrane protein with homology to immunoglobulin 21

Tissue distribution of the Thy-1 determinant 21
Molecular properties of Thy-1 22
Protein structure of Thy-1 23
Homology to Ig and speculation on the function of the Thy-1 antigens 25

References 29

CHAPTER II THE HEAVY CHAIN OF THE HUMAN B CELL ALLOANTIGEN HLA-DS HAS A VARIABLE N-TERMINAL REGION AND A CONSTANT IMMUNOGLOBULIN-LIKE REGION 36

References 49

CHAPTER III THY-1 cDNA SEQUENCE SUGGESTS A NOVEL REGULATORY MECHANISM . 51

References 62

CHAPTER IV THE NUCLEOTIDE SEQUENCE OF THE MOUSE THY-1 GENE 64

Summary 65

Introduction 66

Results 69

Isolation and characterization of mouse Thy-1 genomic clones . . 69
Structure of the mouse Thy-1 gene 74
The gene organization of mouse, rat and human Thy-1 77
Northern blot analysis of RNA from mouse and rat tissues 77
Comparison of 5' untranslated region. 80
Comparison of the predicted amino acid sequences of the Thy-1 genes 85

Discussion 90

Experimental procedures 94

DNA cloning 94
DNA preparation 94
DNA sequencing analysis 95
RNA analysis 96

Acknowledgements 97

References 98

CLOSING STATEMENT 103

LIST OF TABLES

Table Page

Chapter I

1. Amino acid analyses of the Thy-1 glycoproteins 24

Chapter IV

1. Degree of predicted Thy-1 protein sequence homology 88

2. Comparison of amino acid compositions of the Thy-1 antigens 92

LIST OF FIGURES

Figure — Page

Chapter I

1. Genetic map of the MHC in mouse and man 5
2. Structure of class I and class II molecules 5
3. Comparison of published amino acid sequences of class II anigens . 12
4. Exon-intron organization of class II and related genes 15

Chapter II

1. cDNA and predicted amino acid sequence of the DSα subunit clone, pDSα-12 40
2. Comparison of the protein sequences of DR and DSα subunits predicted from cDNA clones 42
3. Comparison of the DNA sequences of DS7 and DS4, 6α subunits . . . 45
4. Comparison of the protein sequences of DS4, 6 and DS7α subunits . 47

Chapter III

1. Partial restriction map and strategy for sequencing Thy-1 cDNA clone, pT64 55
2. Primary structure and predicted amino acid sequence for pT64 . . . 57
3. Comparison of the DNA sequences of mouse $V_\lambda 1$ (MOPC104E) and rat Thy-1 61

Chapter IV

1. Scheme for the use of c2RB vector in the construction of a mouse C57BL/6 cosmid library 71
2. Restriction enzyme maps of Thy-1 genomic clones 73
3. Sequence of the mouse Thy-1.2 gene 76
4. Comparison of the gene organization of mouse, rat and human Thy-1 . 79
5. Northern blot analyses of RNA from mouse rat tissues 82

6. Comparison of the 5' end nucleotide sequences of rat cDNA and mouse and rat genomic DNA 84

7. Comparison of predicted amino acid sequences of mouse, rat and human Thy-1 87

ABBREVATIONS

MHC	major histocompatibility complex
HLA	human leukocyte antigen
Ig	immunoglobulin
Ia	I region associated
β2-M	β2-microglobulin
2-D	two dimensional
L	leader
CP	connecting peptide
TM	transmembrane peptide
CY	cytoplasmic domain
3'UT	3' untranslated region
SDS	sodium dodecyle sulfate
V domain	variable domain
C domain	constant domain
V_L	variable domain of the light chain
V_H	variable domain of the heavy chain
HV	hypervariable
Amp^r	ampicillin resistance
Kan^r	kanamycin resistance
EDTA	(ethylenedinitrilo)-tetraacetic acid
Tris	Tris (hydroxymethyl)aminoethane
α-LA	α-lactoalbumin

INTRODUCTION

This thesis decribes the molecular cloning of two cell surface alloantigens, HLA-DS and Thy-1, both of which are molecules with homology to immunoglobulins.

The HLA-DS alloantigen is a class II alloantigen of the human MHC (Major Histocompatibility Complex). The MHC is a group of closely linked loci coding for molecules that play a critical role in rejection of organ transplants and the control of the immune response. The loci fall into three classes, class I, class II, and class III. The class I and class II molecules, typified by transplantation antigens and the Ia (I-region associated) antigens, respectively, are integral membrane proteins involved in the recognition that permits the immune system to distinguish between self and nonself. The class III family encodes several components involved in the activation pathway of complement. These genes and molecules have been studied intensively over the last five decades by geneticists, biochemists and immunologists, but only recently has the isolation of the genes by molecular biologists facilitated their precise characterization. Many surprising findings have been made concerning their structure, multipilicity, organization, function and evolution. Since my studies relate to the molecular cloning of HLA-DS, I will concentrate my review on the structure, organization and evolution of the MHC class II molecules.

Thy-1 was originally defined in mice as a cell surface alloantigen of thymus and brain with two allelic forms, Thy-1.1 and Thy-1.2. It is

present on the surface of T lymphocytes, but not B lymphocytes, of the mouse, so it has been used as a mouse T cell marker. Subsequently, the Thy-1.1 alloantigenic determinant was identified in rats. In contrast to mice, the rat Thy-1 determinant is present on bone marrow cells, but not T lymphocytes. Thy-1-like molecules have also been found in many species, and in a variety of cells. The tissue distribution of the Thy-1 determinant reveals a very unusual pattern within species, and between species. Recently, the proteins expressing the Thy-1 determinant have been isolated, and their protein sequences have been determined. The homology between Thy-1, Ig, and MHC molecules has been noted. Therefore, I take the liberty to discuss the Thy-1 molecule with the MHC molecules, and will generally discuss the isolation, tissue distribution, relateness of the Thy-1, Ig and MHC molecules and the speculative functions of the Thy-1 molecule.

Chapter I

LITERATURE REVIEW

I. The Major Histocompatibility Complex in mouse and man

General introduction of the MHC. The MHC was discovered by P.A. Gorer in the 1930s (1). It was found that tumors of a particular mouse would grow only in mice of the same inbred strain and that the alloantigen of red blood cells, antigen II, is important in determining the fate of these transplants (2). Subsequently, the genes postulated in the genetic theory of tumor transplantation were referred to as histocompatibility genes, H-2, by G.D. Snell (3). By grafting tumors or skin among such mice and following the rejection or acceptance of the graft, the H-2 was determined to be a large, complex genetic region, and was denoted the Major Histocompatibility Complex, MHC. The cell surface structures responsible for graft rejection were initially characterized using alloantibodies produced by crossimmunization of mice differing only at the MHC. The specificities of such alloantisera for gene products of the murine MHC have permitted the identification of three classes of molecules (4). Class I molecules, typified by transplantation antigens, serve as restricting elements for the response of cytotoxic T cells to virally infected cells. Class II molecules, also called I-region associated, or Ia antigens, regulate the recognition of foreign antigens on the surface of macrophages and B cells by T cells (5). Class III molecules include several components in the activation stage of complement cascade (6), and will not be discussed further.

The H-2 Complex and the Human Leukocyte Antigens (HLA). The genetic maps for the MHC of mouse and man have been constructed by serological analysis of recombinant MHC chromosomes (fig. 1) (4, 5, 7, 8). The H-2 complex is located on chromosome 17, and spans about 2 centimorgans of DNA, which corresponds to approximately 4000 Kb of DNA. The complex is divided into six subregions called K, I, S, D, Qa, and Tla (4,5). Two categories of class I antigens are encoded at four loci of the H-2 complex--K, D, Qa and TLa. The class I loci, K and D, encode cell surface molecules termed transplantation antigens, denoted H-2K, H-2D and H-2L antigens, which mediate the graft rejection initially used to define the MHC. The class I molecules belonging to this category represent the classical H-2 antigens. They are expressed on all tissue types, and are highly polymorphic, with more than 50 different alleles (4,5). The second category of class I genes, Qa and Tla, encode the Qa-1, Qa-2 and Tla antigens, which are structurally closely related to the class I transplantation antigens. However, they differ from classic H-2 antigens in several aspects. The Qa antigens are preferentially expressed on B and T cells, and the Tla antigens on thymocyte and certain leukemia cells. More striking, the Qa and Tla antigens exhibit a much lower level of polymorphism among mice (9). Class II genes lie between the K and D loci, in the region called I of H-2, and encode the cell surface Ia antigens, which are primarily expressed on the surface of B cells, macrophage, and activated T cells, and also exhibit extensive serological polymorphism (4, 5). Two distinct types of murine Ia molecules, I-A and I-E, have been identified (10).

The human MHC called human leukocyte antigens (HLA) complex is located on chromosome 6, and encompasses about 3 centimorgans of DNA,

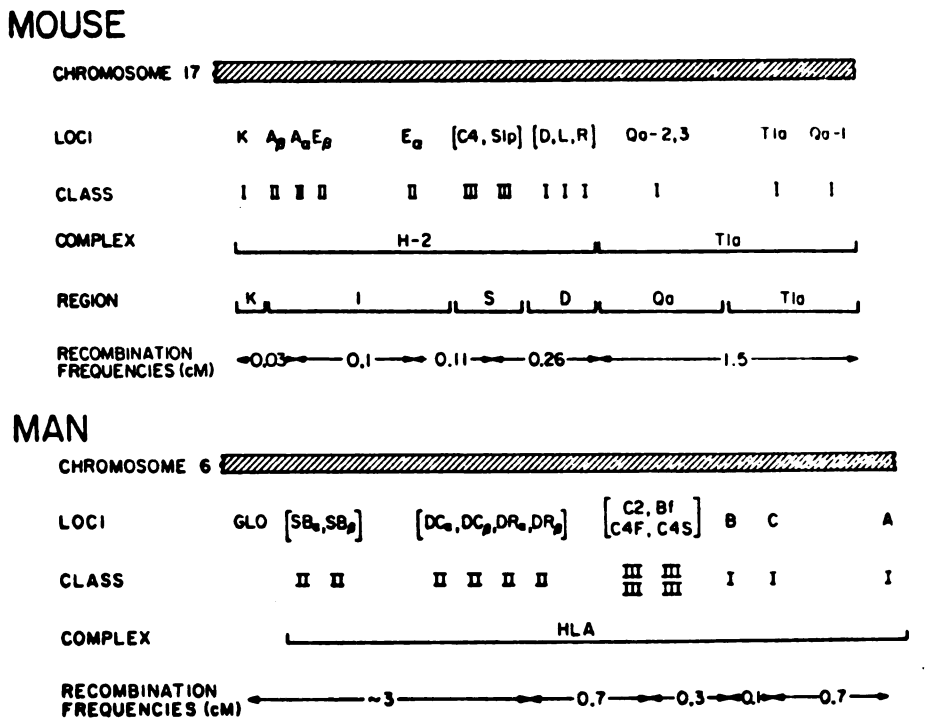

Figure 1 Genetic map of the MHC in mouse and man.

(Taken from reference 8)

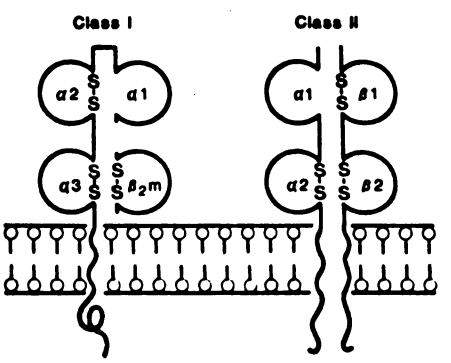

Figure 2 Structure of class I and class II molecules. S-S indicates disulfide bridges in the extracellular domains of class I and class II molecules and β2-microglobulin (β2-M). (It is taken from reference 7).

extending perhaps over 6000 Kb of DNA (fig. 1). It also consists of three types of loci (4, 5, 7). Class I loci HLA-A, -B, -C (analogous to K and D of mouse H-2), encode the transplantation antigens which are present on virtually all cells and elicit strong allograft reactions. The gene organization of human class I loci is different from their mouse counterparts, since the HLA-A, -B, -C genes are contiguous to one another, whereas in the mouse H-2 K gene is separated from the other class I genes by H-2 I and H-2 S loci. The human counterparts to mouse Qa and Tla antigens have been charaterized serologically and biochemically, but their genes have not yet been mapped (11). Class II loci HLA-D (analogous to I of mouse H-2) encodes three distinct types of human class II molecules, DR (12), DC(DS) (13), and SB (14) antigens which stimulate response of T cells in the mixed lymphocyte reaction and probably modify the response to antigen on macrophage.

Structure and function of the MHC antigens. Both class I and class II antigens are heterodimeric cell surface glycoproteins (fig. 2) (7). The class I loci encode for a polypeptide chain with a molecular weight of 45 KD, which contains three extracellular domains ($\alpha 1$, $\alpha 2$ and $\alpha 3$), each about 90 amino acids long, a transmembrane region, and a cytoplasmic domain (4, 5, 7, 8). The third extracellular domain is noncovalently associated with $\beta 2$-microgloblulin ($\beta 2$-M) on the cell membrane. The $\beta 2$-M is a 12 KD polypeptide (15) which shows homology to the constant domain of Ig and is encoded on the murine chromosome 2 and human chromosome 5 (16). Thus the cell surface molecules of class I appear to be heterodimers, but only the class I molecules are anchored on the membrane. Class II antigens consist of two non-covalently associated polypeptide chains with

molecular weight of 33,000 and 28,000, designated heavy chains (α) and light chains (β), respectively. Both of these polypepetides are encoded in the MHC, each contains two extracellular domains, a transmembrane region, and a small cytoplasmic domain (4, 5, 7, 8, 16).

All the cell surface antigens of the MHC have a cruical role in the regulation of the immune response. The class I antigens, originally recognized as transplantation antigens have been shown to be involved in the recognition and killing by cytotoxic T cells, that permit T cells to detect foreign antigens in the context of self (17). In the process of immunosurveillance, the T cell receptor on the cytotoxic T lymphocyte must recognize both foreign antigen and self transplantation antigen, This phenomenon has been termed major histocompatibility restriction, or H-2 restriction in mice. The function of non polymorphic class I antigens, Qa and Tla, is unknown. Class II antigens also function as restricting elements, but for regulatory T cells (18). The proliferative response of T cells to antigen presented on the macrophage surface required histocompatibility at the H-2 I region of the two cell types. Similarly, the cooperation between helper T cells and B cells exhibits H-2 restriction. Thus, the Ia antigens determine several immunologic reactions including control of the level of immune response (proliferation and suppression), delayed-type hypersensitivity (DTH), disease-susceptibility, and primary and secondary allogeneic T cell reactions.

Both the importance of T cell recognition in the positive and negative regulation of the immune system and a number of unusual characteristices of class II molecules, make the studies of the class II molecules important. For instance, the class II molecules possess unusual features of biosynthesis, conformational change, and interaction. Their

structures suggest evolutionary relationships with both class I, and Ig molecules. Like class I antigens, they form a highly polymorphic multigene family with a complex pattern of antigenic determinants. Recently, the rapid progress of molecular biology has dramatically advanced our understanding of the MHC. The isolation of genes has facilitated the precise characterization of the MHC class II molecules. Therefore, I will concentrate the following discussion on the topic of the molecular structure, gene organization, and functional studies of class II antigens derived from molecular cloning.

II. Class II antigens of the MHC

The molecular structure of class II antigens. The initial efforts to purify class II molecules on the basis of reactivity with specific antisera led to the identification of a number of integral membrane glycoproteins composed of two chains, a heavy chain (α) and a light (β) chain (7, 8, 16). In mice, two families of Ia molecules encoded by the I-A and I-E subregions have been identified and characterized. Both I-A and I-E molecules consist of two noncovalently associated polypeptide subunits, $I\text{-}A_\alpha$ and $I\text{-}A_\beta$, $I\text{-}E_\alpha$ and $I\text{-}E_\beta$, respectively. However, the I-A and I-E molecules are structurally distinct based on partial N-terminal sequence and peptide map analysis (10, 19). Futhermore, I-A and I-E molecules display different patterns of structural variability. When allotypic I-E molecules are structurally compared, their small (β) subunits are found to be very different, whereas their large (α) subunits are generally invariant. I-A molecules, in contrast, display structural variability in both subunits when allotypes are compared.

Recent studies have identified three distinct types of human class II molecules. The HLA-DR molecules are structurally homologous to the murine I-E molecules (10, 12), whereas the DC(DS) molecules are the human equivalent to the murine I-A molecules (13). These molecules display the same patterns of structural variability as their murine counterparts. The third human class II molecule, SB, has not been found in the mouse, although the SB gene appears to crosshybridize most strongly with the mouse $E_{\beta}2$ gene (7, 32).

The isolation of genes coding for class II antigens. Due to the availability of specific antisera and recent advances in molecular cloning techniques, the genes coding for class II molecules have recently been isolated. Two procedures were used to isolate human class II cDNA clones. First, cDNA clones were selected for their ability to bind class II mRNA, which could be identified by *in vitro* translation and immunoprecipitation of the synthesized polypeptide. Such an approach was used in the cDNA cloning of DRα (20), DCα(21), DRβ(22) and DCβ(23). In an alternative approach, short oligonucleotides were synthesized corresponding to the nucleotide sequences predicted from amino acid sequences of class II polypeptides, and were used as hybridization probes to screen for the class II cDNA clones (24). The genomic clones containing human class II genes were isolated by hybridization with the cDNA clones (25). Subsequently, the cDNA and genomic clones encoding mouse class II antigens were isolated by cross species hybridization with human cDNA clones (26-32).

Polypeptides of class II antigens. Amino acid sequences deduced from

DNA sequences of cloned genes (19-42) have complemented the N-terminal sequence data and other structural studies in providing a clear picture of the molecular structure of class II antigens. These sequences, together with studies demonstrating selective proteolytic cleavage of the dimeric complex, show that each molecule is composed of four extracellular domains, two N-terminal domains with no discernible homology to Ig ($\alpha 1$ and $\beta 1$), and two immunoglobulin-like domains (closest to the membrane) with sequence similarities to the constant Ig domain ($\alpha 2$ and $\beta 2$). A short hydrophilic peptide connects the extracullar domains to a short membrane-binding hydrophobic region, and continues with a short hydrophilic carboxyl-terminal domain located in cytoplasm (fig. 3).

The first domain of class II heavy chains are variable in length (from 85-88 amino acids), and contain one glycosylation site at homologous residues (residues 82 for $A\alpha$). The sequences of $DR\alpha$ 1 from different alleles are invariant, and homologous to murine E domain. The DC(DS)$\alpha 1$ sequences are variable between different alleles, and show sequence homology to the murine $A\alpha$ domain. The first domain of the class II light chain is 95 amino acids long and has a glycosylation site at residue 19. Two cysteines at positions 15 and 79 form a disulphide loop of 64 amino acids. The human class II $\beta 1$ domains have strong homology with $A\beta 1$ and $E\beta 1$ domains, and low homology with human class I $\alpha 1$ domain, and $\alpha 2$ domain (35, 38).

The econd domains of both chains ($\alpha 2$ and $\beta 2$) are 95 amino acids long, bear a 56 amino acid disulphide loop (from residues 111 to 167 in $\alpha 2$ for $A\alpha$, and from 118 to 174 in β 2), with one glycosylation site in the $\alpha 2$ domains, at residue 122 for A_α. It has been shown to have strong sequence homology to Ig constant regions $C_\gamma 3$ and $C_\mu 4$, the $\alpha 3$ domain of

Figure 3. Comparison of published amino acid sequences of class II antigens. The sequences are divided into three parts accroding to the Aα chain exon structure: (a) the first extracellular domain, (b) the second extracellular domain and (c) connecting peptide, transmembrane and cytoplasmic regions. Known exon boundaries are indicated by arrows. The identical residues are indicated by dashes. The gaps are inserted to improve homology. The glycosylation site triplets are boxed, and the disulfides are indicated by arrows.

Aα : Alleles k, d, b, f, u, q (30). Eα: allele K (27); d (39). DC_α : DR4,6 (21). DS_α : DR7 (33). DR_α : DR4,4 (38); DR maja (40); DR untypied, DR2,2, (41). Aβ : alleles b, d, k, (42). Eβ : allele d (28). DC_β : DR4,6 (35); DR2,2 (Taken from ref. 7.). DR_β : DR4,6 (34); DR2,2 (Taken from ref. 7.).

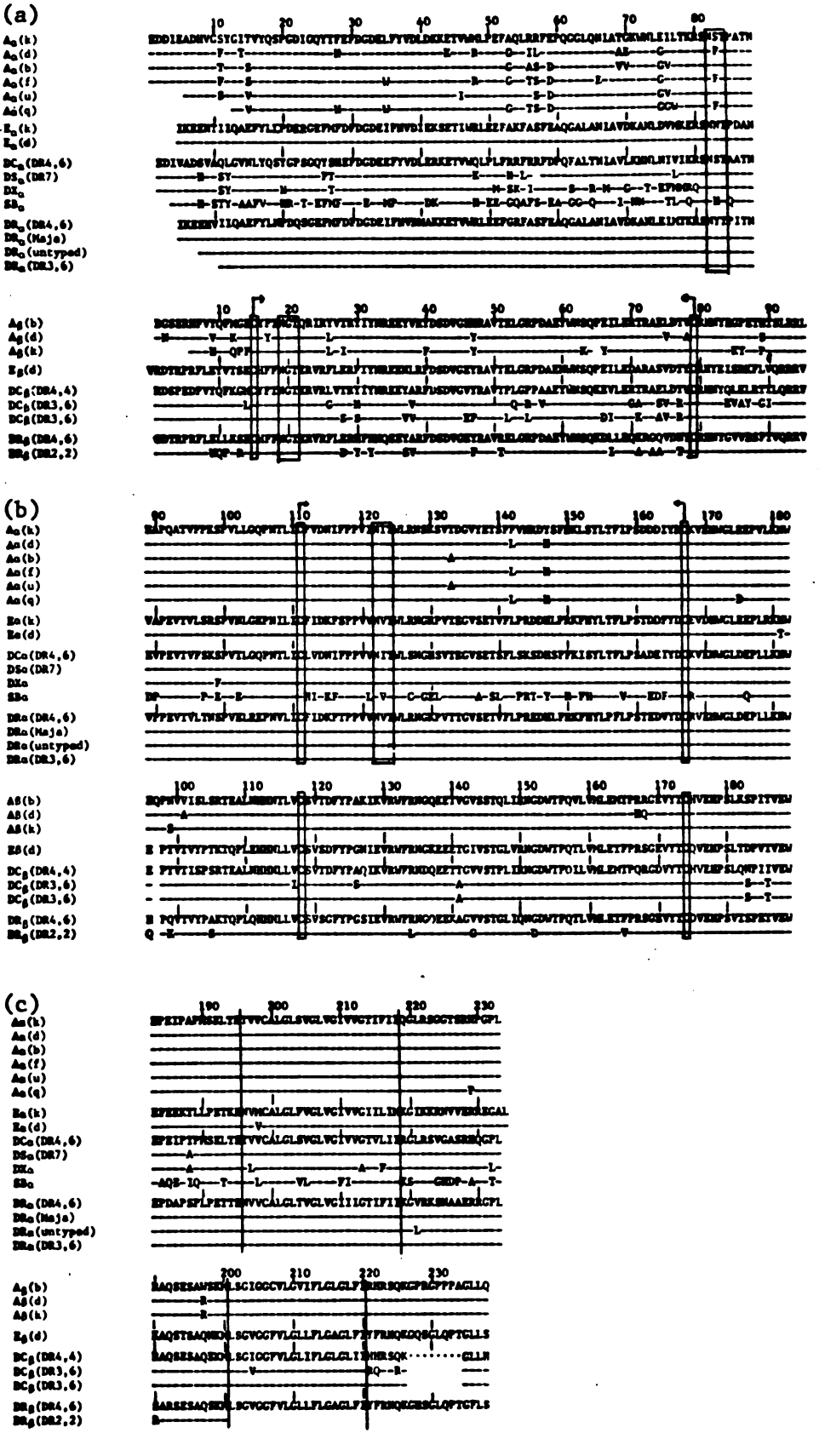

Figure 3

class I antigens, β2-microglobulin and Thy-1 antigens (25, 35, 38).

Following the second domain, there is a hydrophilic peptide called the connecting peptide (CP) that connects the extracellular domain with the transmembrane region. The CP of the heavy chains are 13 amino acids long and rich in glutamic acid and proline, whereas the CP of the light chains are 11 amino acids long and rich in serine. There is no sequence homology between the CP of the heavy chain and light chain.

As expected, there is no detectable homology between the heavy chain and light chain in the transmembrane region. However, there are striking homologies between different heavy chain transmembrane hydrophobic regions, and also between light chain hydrophobic sequences. Strong homologies in the hydrophobic regions of other molecules have been reported, including membrane IgG1, IgG2 and IgM (43), Thy-1 antigens of mouse, rat and human, and glycophorins of man, ox, and pig (44). This indicates that the structure of the hydrophobic domain is conserved to serve an important, specific, but unknown biological function for each molecule. In class II molecules, it has been suggested that interaction between the two chains in the hydrophobic region provides a selective pressure to conserve these homologies. The hydrophobic region ends, as with every transmembrane protein sequenced, with a cluster of positively charged residues that are thought to interact with the negatively charged phospholipid headgroups of the inner leaflet of the membrane. These residues are followed by short cytoplasmic tails of variable length containing many charged and hydrophilic residues. Mouse and human class I and class II genes show similar variations in the length of their cytoplasmic domain. The cytoplasmic tails of membrane bound polypeptides have been proposed to mediate the effects of external stimuli to the cell

interior. It is not clear whether the variations in length of the cytoplasmic domains are important for different effector functions of the molecules and for distinct interactions with cytoskeletal proteins, or whether they simply reflect evolutionary variations of a non functional carboxyl terminus.

Exon and intron organization of class II genes. Genomic cloning of DRα (25), DCβ(8), E_α (27), Eβ(28) and A_β (32) genes have determined the intron-exon organization of these genes. There is a striking correlation between the organization of exons in the genes and the structural domains of the class II molecules--ie., each structural domain is encoded by a separate exon (fig. 4). In the DRα and E_α , the leader peptide (L) and first domain (α1) is separated by a large intron; the transmembrane region, cytoplasmic exon and part of the 3' untranslated region is encoded in a single exon, followed by an intron dividing the 3' untranslated region into two exons (an unusual feature for an eucaryotic gene). A similar gene organization can be found in the gene for β2-M (45). The gene organization of class II light chains, A_β , E_β, and DC_β are different from that of the heavy chains. The β1 and β2 domains are separated by a large intron, 2.4 kb in A_β and DC_β and 3.9 Kb in E_β. The transmembrane region is encoded by a single domain, whereas the cytoplasmic domain is split into two exons and the 3' untranslated region is encoded in a single exon. The gene organization of the class II light chain genes are similar to class I heavy chain genes (8). Thus, the structure of class II α and β genes differ in several regards. The DCβ cytoplasmic region lacks eight amino acids present in the other light chains of I-A and I-E, which correspond to a single exon in the A_β and E_β (28, 32). Whether this reflects

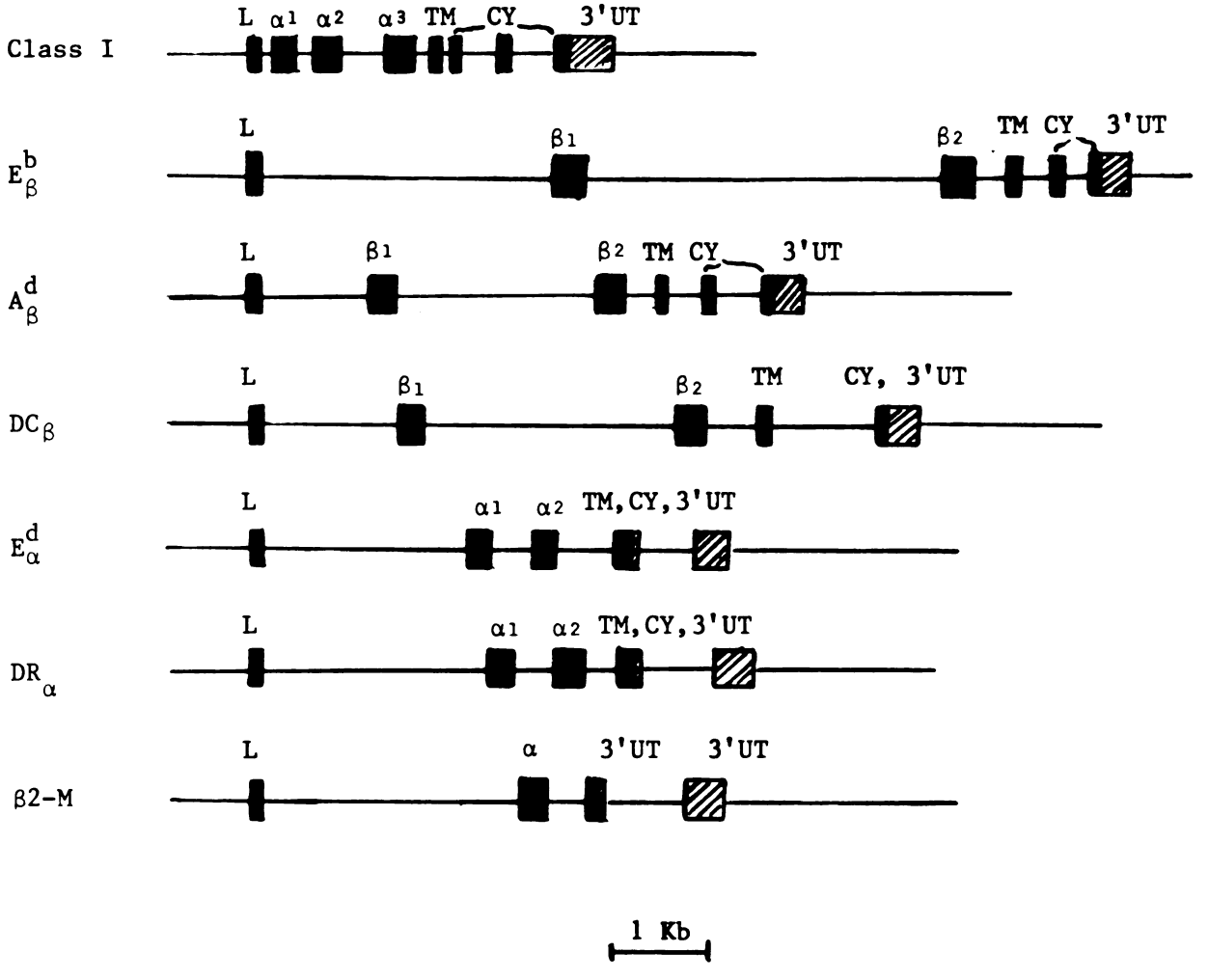

Figure 4. Exon-intron organization of class II and related genes. L, leader peptide; α1, α2, β1, β2, extracellular domains; TM, transmembrane domain; CY, cytoplasmic domain; 3'UT, 3' untranslated region. Data taken from references in Figure 3.

deletion of the exon, loss of a splice site, or alternative splicing events is not clear. One of two DR light chain cDNA clones reveals a lack of all of the transmembrane and cytoplasmic region (16). Whether this represents a secreted form of class II molecules needs to be clarified.

Genetic map of the class II region of the MHC. The genetic map of the genes coding for class II antigens has been divided into five subregions: I-A, I-B, I-J, I-E and I-C, by recombinational analysis (4, 5). Recently, the organization of the class II genes in the I region of the BALB/c mouse has been studied by chromosome walking procedures (26). A human DRα cDNA probe was used to screen a cosmid library. This was followed by using left-ended or right-ended single-copy probes for the sequential screenings, and over 200 Kb of DNA representing most of the I region of BALB/c mice has been isolated. Several class II genes, $A\alpha$, $A\beta$, $E\alpha$, $E\beta$, and two unknown genes or pseudogenes called $A\beta 2$ and $E\beta$ 2, have been mapped in this region (26). The E_α, A_α, and A_β genes were identified by direct DNA sequence analysis, the E_β gene has been identified by hybridization with an oligonucleotide probe specific for the E_β^k polypeptide, and the $E\beta$ 2 gene has been identified by cross hybridization to the mouse $E\beta$ and human DC β cDNA probes (26). The A_β and E_β , together with $A\beta 2$ have the same 5' to 3' orientation, whereas the E_α, A_α genes show an opposite orientation. The location of the A_α gene between the A_β and $E\beta$ genes is different from the gene order of $A\alpha$-$A\beta$-$E\beta$ suggested by the peptide map analysis of A polypeptides from intragenic recombination studies (46). However, the gene cloning data unequivocally shows that the class II gene order in the BALB/c mice is 5'-$A\beta^2$-A_β-A_α-E_β-E_α-$E\beta^2$-3' (16).

The organization of human class II genes is not known to the same

extent as in the mouse, although an intensive effort is being made to understand the organization of the D region, which appears to be more complex than the I region. When DRβ and DCβ cDNA probes are used in the Southern blot analyses of human DR homologous cell lines, 4-13 bands (depending on the restriction enzyme used) can be detected, suggesting the existence of at least three categories of light chain genes in human MHC (47-49). Isolation of cosmid clones, together with mapping using γ-ray induced deletion mutants, show that five or six heavy chain genes are located in the D region (50, 51). It is currently believed that there are at least seven human light chains (three DRβ , two DCβ and two SBβ) and at least five or six human heavy chains (one DRα, three or four DCα-related, and one SBα).

Polymorphism of class II antigens. Serological studies have defined the several alloantigenic systems of the MHC, I-A and I-E antigens in mouse and DR, DC(DS), and SB antigens in human. Each of these alloantigenic systems can be distinguished from the others on the basis of immunodepletion and two-dimensional (2-D) gel analyses. These systems are all highly polymorphic. By using alloantisera and mixed lymphocyte reactions, at least 13 DR alleles can be defined (52). Biochemical studies, including 2-D polyacrylamide gel electrophoresis, tryptic peptide analysis, and N-terminal protein sequence determination, have suggested that the polymorphism of class II antigens results from their structural differences. In most biochemical studies, the β chains have been cited as being divergent, while very limited polymorphism has been attributed to the α chains (10). The amino acid sequences derived from cDNA clones of class II antigens have led to the conclusion that alleles of polymorphic

class II loci are very different from each other (Figure 3). The differences are found not only on the light chains but also on the heavy chains of A and DC.

The comparison of DR α chains has revealed that the DRα chain is invariant, or nearly so, since only one amino acid difference (Val/Leu in the cytoplasmic tail) has been found in several DR$_\alpha$ clones. The E$_\alpha$ chain sequences from alleles d and k differ only by two amino acids (one in the α 2 domain, the other one in the transmembrane region). However, the heavy chains of the I-A and DC antigens show more variation between different alleles. Comparison of Aα sequences from several alleles shows 7-18 amino acid differences, whereas 11 amino acid differences can be found between the DCα chains of DR7 and DR4,w6 cell lines (33). Most of these differences are concentrated in the α1 domain and are clustered within a few hypervariable regions (31, 33). When the amino acid sequences of class II light chains are compared between different alleles, they reveal 13-15 amino acid differences between different Aβ alleles, 17-27 amino acid differences between DCβ chains, and 24 amino acid differences between DRβ chains of DR4, 6 and DR2, 2 cell lines (fig. 3). The overall sequence of the α chain is less ploymorphic than that of the β chain. However, the β1 domains are not substantially more polymorphic than α1 domain of the I-A and DC(DS).

The molecular mechanisms involved in the generation of the extensive allelic polymorphism of class II antigens have not been established. Several hypotheses have been proposed. An analysis of nine intra-I region recombinants has demonstrated that all nine recombination events have occured within a distance of 8 Kb or less at the Eβ locus--ie., a hot spot of recombination (26). Together with the data from studies of determinant

shuffling, this suggests that, as for class I antigens (53), recombination and gene conversion between members of the class II multigene family are important mechanisms for generating diversity (26). However, another hypothesis has been suggested to explain the generation of Aα allelic polymorphism (32).

Biosynthesis of class II antigens. As mentioned above, the class II molecules have a very unusual feature of biosynthesis. It has been reported that the biosynthetically immature class II antigens are associated with an invariant third chain, provisionally called In (Ii,γ) chain (54-56). More specifically, newly synthesized α and β chains form a complex with the In chain in the endoplasmic reticulum. After transport of the protein complex to the Golgi apparatus, and concomitant with terminal glycosylation, the In chain dissociates from the α and β chains. At least a fraction of the In chains subsequently become integrated into the plasma membrane independently from class II antigens. The In chain is a basic, methionine rich, transmembrane glycoprotein, and is not encoded within the MHC (55). The biological function of the In chain is not yet known. It has been suggested that it may regulate the intracellular transport of class II antigens, and it may prevent the formation of class II hybrid molecules-- ie., molecules composed of α and β subunits encoded by different loci. The cDNA clone of a human In chain has been isolated (56). The amino acid sequence derived from DNA analysis suggests that the N-terminus of the In chain resides on the cytoplasmic side of membrane, and the In chain may be devoid of an N-terminal signal. DNA mediated gene transfer experiments of DR or I-A genes to mouse L cells, which contain very little In chain, indicate that the In chain may not be

essential for the surface expression of class II molecules. The precise role of the In chain still remains to be clarifed.

Gene expression and functional studies of class II genes. The class II molecules have long been proposed to serve as restricting elements that permit the regulatory T cell (helper, suppressor, and amplifier) to view antigen in the context of self on the surface of other T cells, macrophages, or B cells. Class II genes appear to control the proliferation of the regulatory T cells as well as the effector reactions carried out by these T cells, such as the amplification of other T-cell subsets and the promotion of B-cell differentiation. The isolation of genes, gene transfer and in vitro site-directed mutagenesis techniques offer complementary approaches to dissect the function of each molecule, and to clarify the structure-function relationships existing among Ia molecules, foreign antigens and T cell receptor molecules. Malissen et al. have transferred the mouse Aα and Aβ genes into mouse L-cell fibroblasts and hamster B cells (57). The I-A molecules expressed on these two types of cells appear normal by serological assays and 2-D gel electrophoresis, and can present certain antigens to T cell helper hydridomas. Futhermore, the I-A molecules on the L cell can act as targets for allogenic cytotxic T cells. Rabourdin-Combe and Mach also have successfully transferred the DRα , DRβ and the DR associated In chains into mouse L cells (58). The molecules expressed have been identified by monoclonal antibodies. Whether the molecules also possess the function of DR antigens is still unknown.

III. The Thy-1 antigen--a membrane protein with homology to immunoglobulin

Tissue distribution of the Thy-1 determinant. The Thy-1 antigen was first identified in the mouse by A. E. Reif and J. M. Allen (59) as a cell surface alloantigen of thymus and brain, with two allotypic forms called θ-AKR (Thy-1.1) and θ-C3H (Thy-1.2). Both antigenic determinants are coded by the Thy-1 locus on chromosome 9 in mice (60). By using mouse alloantisera, the Thy-1.1 but not the Thy-1.2 alloantigenic determinant was recognized on rat thymocytes and brain cells (61). The mouse and rat Thy-1.1 determinants are similar but not identical, since the affinity of cross-reacting antibody differs by 10-fold between two species. Other antigenic determinants on the Thy-1 molecule can be recognized by xenogeneic antisera, and these include species-specific and cross-reacting determinants (62). The Thy-1 antigens defined either by allo- or xeno-antisera have been shown to be located on the same molecular complex, which is referred to as the Thy-1 molecule. The Thy-1 antigens have also been identified in dog and man by cross-reaction with rabbit antibodies to rat Thy-1 (63).

The tissue distribution of the Thy-1 antigen follows unusual patterns within a species, and shows surprising differences between species (63). In all species studied the molecule is a major constituent of the brain cell membrane, where there are about 600,000 molecules per cell (64). It is predominantly found on neuronal cells. In lymphoid tissues the Thy-1 antigen is likely to be the most abundant cell surface molecule of mouse and rat thymocytes, but is found in reduced amounts in dog thymocytes and not at all in human thymocytes. In the mouse, Thy-1 is absent from stem cells, and has been widely used as a marker for T lymphocytes. In contrast

to this, most rat T lymphocytes lack Thy-1, but the molecule is found on a sub-set of bone marrow cells (30-45%), including immature B cell and haemopoietic stem cells (65). In rodents some fibroblasts (66), epidermal cells (67), breast cells (68), and muscle cells (69) also display the Thy-1 antigen. In many of these tissues the levels of Thy-1 expression undergoes dramatic changes during differentiation (68, 69).

Molecular properties of Thy-1. All of the Thy-1 antigenic determinants from the thymus and brain of rats and mice are expressed in glycoproteins with an apparent molecular weight on SDS-polyacrylamide gels of about 25 KD. The molecules have been purified from both tissues in both species (70, 71). The molecular weights, determined by sedimentation equilibrium measurements, are 17500 for brain Thy-1 and 18700 for thymocyte Thy-1, and in each case the molecular weight of the polypeptide is 12500. The amino acid compositions of the Thy-1 glycoprotein from brain and thymus are very similar (72), but the carbohydrate composition of brain Thy-1 glycoprotein differs from that of thymocytes, accounting for the difference in molecular weight (72).

Both thymocyte and brain Thy-1 glycoproteins have the properties of molecules that bind directly to the plasma membrane. Rat Thy-1 glycoprotein can be solubilized as a monomer in deoxycholate (DOC) and detergent Brij 96, and binds one micelle of DOC/molecule. If the detergent is removed, the molecule self-associates to form an oligomer. Mouse thymocyte Thy-1 glycoprotein is labelled in membranes by a lipophilic, photoactivatable reagent which labels only the hydrophobic section of membrane proteins (73). These properties suggest that the molecule has a hydrophobic portion in its structure and this is presumably responsible

for integration into the membrane since there has never been any indication of Thy-1 glycoprotein binding to another molecule in immunoprecipitation studies (74).

Protein structures of Thy-1. The molecules expressing the Thy-1 determinants have been isolated from the brain of mouse (81), rat (70), dog (75), human (76, 77), chicken (78), frog (79) and squid (80). These glycoproteins all have a structure related to the mouse and rat Thy-1 molecules. The amino acid compositions of the Thy-1 molecules from brains of chicken, mice, rats, men and squids, and from human foreskin have been determined (76, 78, 80, 81) (Table 1). The protein sequences of the Thy-1 membrane glycoprotein of mouse and rat brain have been reported.

The rat brain Thy-1 sequence was determined from tryptic and V-8 proteinase peptides, and consists of 111 amino acids. The molecule contains two disulphide bonds, Cys-9 to Cys-111 and Cys-19 to Cys-85, and three N-linked amino acid sugars, located at Asn-23, -74, and -98 (81). The C-terminal peptides were unusual, in that they were either obtained in a highly aggregated form, or could only be purified after binding to detergent Brij 96 micelles. They thus appeared to have hydrophobic properties, yet did not contain any extended sequence of hydrophobic amino acids. The C-terminal peptides also contained some unidentified ninhydrin-postive material, glucosamine and galactosamine. This suggested that the hydrophobic properties of the C-terminal peptides may be due to the linkage of lipid.

Mouse Thy-1 glycoproteins were purified from the brain of AKR strain (Thy-1.1) and CBA strain (Thy-1.2) mice, and the protein sequences were determined (80). The mouse sequences of mouse Thy-1 are very similar to

Amino acid analyses of the Thy-1 glycoproteins

	Human Brain Thy-1	Rat Brain Thy-1	Mouse Brain Thy-1.2	Chicken Brain Glycoprotein	Squid Brain Glycoprotein
CYS	5	4	4	4	8
ASX	9	15	14	14	15
THR	11	10	10	10	7
SER	12	8	11	12	7
GLX	11	10	11	10	12
PRO	5	3	3	3	3
GLY	5	4	4	6	8
ALA	4	2	3	5	5
VAL	8	9	8	7	8
MET	1	1	1	2	2
ILE	2	5	4	6	5
LEU	11	13	12	12	6
TYR	5	2	4	5	4
PHE	4	2	4	3	4
HIS	5	4	4	3	4
LYS	8	8	9	8	9
ARG	6	9	7	6	6

Data were taken from references: human Thy-1 (76); mouse Thy-1.2 (80); rat Thy-1 (81); chicken Thy-1-like glycoprotein (78); squid brain glycoprotein (80).

that of rat Thy-1. Both molecules consist of 112 amino acids, with the two allotypic forms differing by only one amino acid at residue 89, where Thy-1.1 has arginine and Thy-1.2 has glutamine. This fits well with the amino acid composition, which is similar except that Thy-1.1 has one more Arg and one less Gln than that of Thy-1.2 (76). It also explains the finding that Thy-1.1 is more basic than Thy-1.2 in isoelectrophoresis (82).

Homology to Ig and speculation on the function of the Thy-1 antigens. Homology between Thy-1, which has a domain-like structure including a disulphide loop of the appropriate size, and the C domain of Ig has been noted at the protein sequence level by E. Cohen et al.(83). With the mouse protein sequences and further analysis, it is now clear that the Thy-1 sequence fits better along the whole sequence of the Ig V domain. This was especially suggested by a comparison of the protein sequence of mouse Ig $V_{\lambda}1$ and rat Thy-1 (81), and analysis of identities between the protein sequences of Thy-1 and V_L and V_H (80). There are 25 and 23 sequence identities between mouse Thy-1 and V_L and V_H domain, respectively, and 9 of these in each case are the residues that are conserved in more than 90 % of all V_L and V_H sequences. There are a number of long stretch of identical sequence among the mouse, rat and human Thy-1 molecules. These are interrupted in three places by blocks of residues that differ among the species, residues 23 to 29, 60 to 68, and 99 to 103. The position of these blocks of divergent sequences correspond to the locations of hypervariable regions of the Ig V domain (84). These data suggest that the Thy-1 molecule has homology to both the C and V domains of Ig.

The comparison of the protein sequences of Thy-1 and β 2- M suggests

that the homologies are at least as good as those between β2-M and C domains of Ig (85). DNA and amino acid sequence analyses have revealed homology among the MHC antigens, β2-M and Ig domains, and suggest that all three have evolved from a common ancestral gene encoding a primitive domain (85-87). On the basis of sequence homologies, it was suggested that Thy-1 could be the primordial domain of the Ig superfamily. One homolog of the Thy-1 molecule with the similar amino acid composition to the other Thy-1 molecules was found in the squid brain (80). A partial amino acid sequence showed homology with rodent Thy-1 and $Ig_{\lambda}V$ domain sequences, suggesting that this molecule may be the first invertebrate member of the Ig superfamily. This finding greatly strengthened the idea that the Thy-1 molecule is primitive, and is likely to be the primordial domain of Ig family.

Despite the fact that Thy-1 is so abundant on the surface of thymus and brain cells, and that its molecular structure is so well characterized, the function of Thy-1 is still an open question. Because of sequence homology to Ig, it has been suggested that molecules displaying the Thy-1 antigen play a role in cellular recognition and morphogenesis in the nervous system (81). The expression of the Thy-1 antigen accompanying tissue differentiation has been noted since Thy-1 was first discovered. Reif and Allen noted that the expression of Thy-1 in the brain accompanies the development of brain, ie. neonatal brain possesed only 1.5-2.8% of the content of antigen found in adult brain. In contrast, the antigen was almost fully developed in the neonatal thymus (59). In addition, J. E. Lesley and V. A. Lennon also observed that the expression of the Thy-1 antigen on skeletal muscle was transitory, being expressed on myoblasts and newly developed myotubes, but disappearing as myotubes

differentiated. This finding suggests that Thy-1 may be functional in normal muscle development (69).

Recently, many studies have tried to use monoclonal antibodies against the Thy-1 antigen to define the function of Thy-1. Several lines of data have been reported. R. Dulbecco et al have used monoclonal anti-Thy-1 sera to prevent the differentiation of mammary cells (68). Certain monoclonal antibodies to the Thy-1.1 antigen prevent formation of new domes and cause disappearance of preexisting one. This suggests that the specific interaction of these antisera with the Thy-1 antigen redirects the differentiation programs of these cells. A role for Thy-1 in cell interactions leading to differentiation within the thymus, and in enhancing the regeneration of processes by rat retinal ganglion cells, have been suggested (88, 89). The function of Thy-1 in the immune system has also been studied by using either allogeneic antisera or monoclonal antibodies. By using rabbit anti-mouse brain antisera, the Thy-1 antigens were suggested to be involved in T cell activation, since antibodies against Thy-1 stimulate T cell proliferation and induce T cell growth factor production in normal T cells (90). One such monoclonal antiserum which is able to induce interleukin-2 in T cells, was suggested to recognize the Thy-1 molecule (91). Therefore, the Thy-1 antigen has been proposed to be equivalent to human T3 antigen, which has been shown to play a critical role in the T cell antigen recognition process. However, another line of data suggests that the mouse antigen equivalent to human T3 antigen is not the Thy-1 molecule, although it is a 25 KD glycoprotein with the properties very similar to the Thy-1 molecule (92). Whether the Thy-1 antigen is involved in T cell receptor recognition remains to be clarified.

Since the amino acid sequence of Thy-1 has been reported, it is possible to use recombinant DNA methods to isolate Thy-1 gene, and study its genomic structure and its relatedness to the immunoglobulins. Furthermore, by using the isolated gene, we are able to study the expression and function of Thy-1 in various cell types by gene transfer and *in vitro* site-directed mutagenesis techniques.

REFERENCES

1. Gorer, P. A. (1937) J. Pathol. Bacteriol. 44, 691-697.
2. Gorer, P. A. (1938) J. Pathol. Bacteriol. 47, 231-252.
3. Snell, G. D. (1948) J. Genetics 49, 87-103.
4. Klein, J. (1975) Biology of the mouse histocompatibility-2 complex. pp. 192-230. (New York-Springer-Varlag)
5. Snell, G. D., Dausset, J. and Nathenson, S. (1976) Histocompatibility New York Academic Press.
6. Bianco, C. and Nussenzweig, V. (1977) Contemp. Top. Mol. Immunol. 6, 145-176.
7. Steinmetz, M. and Hood, L. (1983) Science 222, 727-733.
8. Hood, L., Steinmetz, M. and Malissen, B. (1983) Ann. Rev. Immunol. 1, 529-568.
9. Yokoyama, K., Stockert, E., Old, L. J. and Nathenson, S. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 7078-7082.
10. Cullen, S. E., Freed, J. H. and Nathenson, S. G. (1976) Transplantation Rev. 30, 236-270.
11. Cotner, T., Mashimo, H., Kung, P. C., Goldstein, G. and Strominger, J. L. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 3858-3862.
12. Ferrone, S. Pellegrino, M. A. and Reisfeld, R. A. (1977) J. Immun. 118, 1036-1041.
13. Goyert, S. M. and Silver, J. (1981) Nature 294, 266-286.
14. Hurley, C. K., Shaw, S., Nadler, L., Schlossman, S. and Carpa, J. D. (1982) J. Exp. Med. 156, 1557-1562.

15. Cunningham, B. A., Wang, J. L., Berggard, I. and Peterson, P. A. (1973) Biochemistry 12, 4811-4821.

16. Kaufman, J. F., Auffray, C., Korman, A. J., Shackelford, D. A. and Strominger, J. L. (1984) Cell 36, 1-13.

17. Zinkernagel, R. M. and Doherty, P. C. (1980) Adv. Immunol. 27, 51-177.

18. Klein, J. and Figueroa, F. (1981) Immunol. Rev. 60, 23-57.

19. Cook, R. G., Siegelman, M. H., Capra, J. D., Uhr, J. W. and Vietta, E. S. (1979) J. Immunol. 122, 2232-2230.

20. Lee, J. S. Trowsdale, J. and Bodmer, W. F. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 545-549.

21. Auffray, C., Korman, A. J., Roux-Dosseto, M., Bono, R. and Strominger, J. L. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 6337-6341.

22. Long, E. O., Wake, C. T., Strubin, M., Gross, N., Accolla, R. S., Carrel, S. and Mach, B. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 7465-7469.

23. Wiman, K., Larhammar, D., Claesson, K., Gustafsson, K., Schenning, L. Bill, P., Bohme, J., Denaro, M., Dobberstein, B., Hammerling, U., Kvist, S., Servenius, B., Sundelin, J.. Peterson, P. A. and Rask, L. (1982) Proc. Natl. Acad. Sci. U.S.A. 79. 1703-1707.

24. Stetler, D., Das, H., Nunberg, J. H., Saiki, R., Sheng-Dong, R., Mullis, K. B., Weissman, S. M. and Erlich H. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 5966-5970.

25. Korman, A. J., Auffray, C., Schamboeck, A. and Strominger, J. L. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 6013-6017.

26. Steinmetz, M., Minard, K., Horvath, S., McNicholas, J., Srelinger, J., Wake, C., Long, E., Mach, B. and Hood, L. (1982) Nature 300, 35-42.

27. Mathis, D. J., Benoist, C. O., Williams, V. E., Kanter, M. R. and McDevitt, H. O. (1983) Cell 32, 745-754.

28. Saito, H., Maki, R. A., Clayton, L. K. and Tonegawa, S. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 5520-5524.

29. Mengle-Gaw, L. and McDevitt, H. O. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 7621-7625.

30. Benoist, C. O., Mathis, D. J., Kanter, M. R., Williams, V. E. and McDevitt, H. O. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 534-538.

31. Benoist, C. O., Mathis, D. J., Kanter, M. R., Williams, V. E., and McDevitt, H. O. (1983) Cell 34, 169-177.

32. Larhammar, D., Hammerling, U., Denaro, M., Lund. T., Flavell, R. A., Rask, L. and Peterson, P. A. (1983) Cell 34, 179-188.

33. Chang, H-C., Moriuchi, T. and Silver, J. (1983) Nature 305, 813-815.

34. Long, O. E., Wake, C. T., Gorski, J. and Mach, B. (1983) EMBO J. 2, 389-394.

35. Larhammar, D., Schenning, L., Gustafsson, K., Wiman, K., Claesson, L., Rask, L. and Peterson, P. A. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 3687-3691.

36. Gustafsson, K., Emmoth, E., Widmark, E., Bohme, J., Peterson, P. A. and Rask, L. (1984) Nature 309, 76-78.

37. Roux-Dosseto, M., Auffray, C., Lillie, J. W., Boss, J. M., Cohen, D. DeMars, R., Mawas, C., Seidman, J. G. and Strominger, J. L. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 6036-6040.

38. Larhammar, D., Gustafsson, K., Claesson, L., Bill, P., Wiman, K., Schenning, L., Sundelin, J., Widmark, E., Peterson, P. A. and Rask, L. (1982) Cell 30, 153-161.

39. McNicholas, J., Steinmetz, M., Hunkapiller, T., Jones, P. and Hood, L. (1982) Science 218, 1229-1234.

40. Lee, J. S., Trowsdale, J., Travers, P. J., Carey, J., Grosveld, F., Jenkins, J. and Bodmer, W. F. (1982) Nature 299, 750-752.

41. Das, K. H., Lawence, S. K. and Weissman, S. M. (1983) Proc. Natl. Acad. Sci. U.S.A. 80. 3543-3548.

42. Choi, E., McIntyre, K., Germain, R. N. and Seidman, J. G. (1983) Sceince, 221, 283-287.

43. Tyler, B. M., Cowman, A. F., Gerondakis, S. D., Adams, J. M., and Bernard, O. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 2008-2012.

44. Murayama, J-I., Tomita, M. and Hamada, A. (1982) J. Membrane Biol. 64, 205-215.

45. Parnes, J. R. and Seidman, J. G. (1982) Cell 29, 661-669.

46. Rose, S. M. and Cullen, S. E. (1981) J. Immunol. 127, 1472-1477.

47. Bohme, J., Owerbach, D., Denaro, M., Lernmark, A., Peterson, P. A. and Rask, L. (1983) Nature 301, 82-84.

48. Owerbach, D., Lernmark, A., Rask, L., Peterson, P. A., Platz, P. and Svejgaard, A. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 3758-3761.

49. Wake, C. T., Long, E. O. and Mach, B. (1982) Nature 300, 372-374.

50. Trowsdale, J., Lee, J., Carey, J., Grosveld, F., Bodmer, J. and Bodmer, W. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 1972-1976.

51. Auffray, C., Kuo, J., DeMars, R., and Strominger, J. L. (1983) Nature 304, 174-177.

52. Bodmer, J. G. (1978) Br. Med. Bull. 34, 233-240.

53. Weiss, E. H., Mellor, A., Golden, L., Fahrner, K., Simpson, E., Hurst, J. and Flavell, R. A. (1983) Nature 301, 671-674.

54. Day, C. E. and Jones, P. P. (1983) Nature 302, 157-159.

55. Kvist, S., Wiman, K., Claesson, L., Peterson, P. A. and Dobberstein, B. (1982) Cell 29, 61-69.

56. Claesson, L., Larhammar, D., Rask, L. and Peterson, P. A. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 7395-7399.

57. Malissen, B., Price, M. P., Goverman, J. M., McMillan, M., White, J., Kappler, J., Marrack, P., Pierres, A., Pierres, M. and Hood, L. (1984) Cell 36, 319-327.

58. Rabourdin-Combe, C. and Mach, B. (1983) Nature 303, 670-674.

59. Reif, A. E. and Allen, J. M. (1964) J. Exp. Med. 120, 413-433.

60. Blankenhorn, E. P. and Douglas, T. C. (1971) J. Hered. 230, 259-263.

61. Douglas, T. C. (1972) J. Exp. Med. 136, 1054-1062.

62. Williams, A. F. (1976) Eur. J. Immunol. 6, 526-528.

63. Dalchau, R. and Fabre, J. W. (1979) J. Exp. Med. 149, 576-582.

64. Acton, R. T., Morris, R. J. and Williams, A. F. (1974) Eur. J. Immunol. 4, 598-602.

65. Hunt, S.V., Mason, D. W. and Williams, A. F. (1977) Eur. J. Immunol. 7, 817-823.

66. Stern, P. L. (1973) Nature (London) New Biol. 246, 76-78.

67. Scheid, M., Boyse, E. A., Carswell, E. A. and Old, L. J. (1972) J. Exp. Med. 135, 938-955.

68. Dulbecco, R., Bologna, M. and Unger M. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 1848-1852.

69. Lesley, J. F. and Lennon, V. A. (1977) Nature 268, 163-165.

70. Kuchel, P. W., Campbell, D. G., Barclay, A. N. and Williams, A. F. (1978) Biochem. J. 169, 411-417.

71. Trowbridge, I. S., Weissman, I. L. and Bevan, M. J. (1975) Nature 256, 652-654.

72. Barclay, A. N., Letarte-Muirhead, M., Williams, A. F. and Faulkes, R. (1976) Nature 263, 563-567.

73. Owen, M., Knott, J. C. A. and Crumpton, M. J. (1980) Biochemistry 19, 3092-3099.

74. Ledbetter, J. A. and Herzenbrg, L. A. (1979) Immunol. Rev. 47, 63-90.

75. McKenzie, J. L. and Fabre, J. W. (1981) Transplantation 31, 275-282.

76. Cotmore, S. F., Crowhurst, S. A. and Waterfield, M. D. (1981) Eur. J. Immunol. 11, 597-603.

77. Ades, E. W., Zwerner, R. K., Acton, R. T. and Balch C. M. (1980) J. Exp. Med. 151, 400-406.

78. Rostas, J. A. P., Shevenan, T. A., Sinclair C. M. and Jeffrey, P. L. (1983) Biochem. J. 213, 143-152.

79. Mansour, M. H. and Cooper, E. L. (1984) J. Immunol. 132, 2515-2523.

80. Williams, A. F. and Gagnon, J. (1982) Science 216, 696-703.

81. Campbell, D. G., Gagnon, J., Reid, K. B. M. and Williams, A. F. (1981) Biochem. J. 195, 15-30.

82. Hoessli, D., Bron, C. and Pink, J. R. L. (1980) Nature 283, 576-578.

83. Cohen, F. E., Novotny, J., Sternberg, J. E., Campbell, D. G. and Williams, A. F. (1981) Biochem. J. 195, 31-40.

84. Bernard, O., Hozumi, N. and Tonegawa, S. (1978) Cell 15, 1133-1144.

85. Peterson, P. A., Cunningham, B. A., Berggard, I. and Edelman, G. M. (1972) Proc. Natl. Acad. Sci. U.S.A. 69, 1697-1701.

86. Steinmetz, M., Frelinger, J. G., Fisher, D., Hunkapiller, T., Pereira, D., Weissman, S. M., Uehara, H., Nathenson, S. and Hood, L. (1981) Cell 24, 125-134.

87. Larhammar, D., Schenning, L., Gustafsson, K., Wiman, K., Claesson, L., Rask, L. and Peterson, P. A. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 3687-3691.

88. Raelder, A., Arndt, R., Raedler, E., Jablonski, D. and Thiele, H. G. (1978) Eur. J. Immunol. 8, 728-730.

89. Leifer, D., Lipton, S. A., Barnstable, C. J. and Masland R. H. (1984) Nature 224, 303-306.

90. Norcross, M. A., Smith, R. T. and Shimizu, S. (1984) J. Immunol. 132, 833-838.

91. Gunter, K. C., Malek, T. R. and Shevach, E. M. (1984) J. Exp. Med. 159, 716-730.

92. Allison, J. P., McIntyre, B. W. Ridge, L. L., Gross-Pelose, J. and Lainier, L. L. (1984) Fred. Proc. 43, 1593.

Chapter II

The heavy chain of the human B cell alloantigen HLA-DS has a variable N-terminal region and a constant immunoglobulin-like region

by

Hsiu-Ching Chang

Tetsuya Moriuchi

Jack Silver

Department of Microbiology and Public Health

Michigan State University

E. Lansing, MI 48824

Reprinted by permission from Nature 305, 813-815.

The HLA-D region of the major histocompatibility complex (MHC) of man encodes polymorphic glycoproteins found predominantly on the cell surfaces of B cells and macrophages. These proteins mediate interactions, required for the induction of immune responses, among cells of the immune system and consequently are referred to as Ia (immune-response associated). Two families of Ia molecules, DR and DS (also known as DC), have been defined, the former analogous to the I-E (ref. 1) and the latter to the I-A molecules of murine MHC (2-4). Both DR and DS molecules consist of two non-covalently associated polypeptide chains with molecular weight 33,000 and 28,000 designated α and β , respectively. The polymorphism of DR molecules is due to structural variation in the small subunit, DR_β, with the large subunit, DR_α, being constant in structure (5-7). In contrast, both subunits, DS_α and DS_β, are structurally variable when DS allotypes are compared (3). We have now isolated a cDNA clone from a DR7 cell line that contains the entire coding sequence for the DSα subunit and have compared its predicted amino acid sequence with that previously deduced from a DSα cDNA clone isolated from a DR4,w6 cell line (8). This comparison reveals that 10 of 11 amino acid differences are located within the α1 (N-terminal) domain and that the α2 or immunoglobulin-like domains are identical.

Poly(A)-containing RNA was isolated from a consanguineous homozygous DR7 cell line (LG-10) and cDNA was synthesized by reverse transcription using an oligo(dT) primer followed by second-strand synthesis using DNA polymerase. A cloned cDNA libray was constructed by insertion of the cDNA into the Pst I site of pBR322 using the poly(dG).poly(dC) homopolymeric extension method (9). Colonies containing cDNA were screened with a nick-translated cDNA probe, pDCH1 (provided by C. Auffray), which encodes the

entire DS(DC) α subunit except the leader sequence and 17 amino acids of the N-terminus (8). From an initial screening of 10,000 colonies, four strongly hydridizing clones were detected, all of which had similar restriction patterns (data not shown); and the one with the largest insert, pDS_{α}-12, was subjected to DNA sequence analysis (Fig. 1). The cDNA insert of pDS_{α}-12 consists of 921 nucleotides. Starting with the methionine initiation codon, ATG, at nucleotide positions 20-22, there is an open reading frame of 762 nucleotides encoding 254 amino acids and terminating with the TGA codon at nucleotide positions 782-784. This is followed by a 3' untranslated region of 123 nucleotides and a poly(A) tail of 14 nucleotides. Nineteen bases upstream of the poly(A) tail is the canonical polyadenylation signal AATAAA. The first 23 codons, amino acids -23 to -1, encode the hydrophobic leader sequence. The amino acids predicted from the next 25 codons are in nearly perfect agreement with the previously determined N-terminal protein sequence of a DS7 alpha subunit (3). By convention we have given the DS allotype the same numerical designation as the DR allotype of the cell line from which it was isolated. Accordingly, DS molecules isolated from a DR7 cell line are designated DS7. The two discrepancies are probably due to protein sequencing errors although the existence of a second DS7 alpha-like subunit cannot be ruled out. Thus, pDS_{α} -12 contains the entire coding sequence for the DS7α subunit.

As has been shown previously (8), the DS(DC) α subunit displays strong amino acid and DNA sequence homology to the DRα subunit (Fig. 2). However, despite the overall high degree of homology between DRα and DS_{α} (54%) there are several regions of very low homology, notably the first 51 amino acids of the coding region, (including the leader sequence) where only 12

Fig. 1. cDNA and predicted amino acid sequence of the DS_{α} subunit clone, pDS_{α}-12. The nucleotide sequence of the coding strand and the predicted protein sequence (single letter amino acid code) are shown. The nucleotide sequence immediately follows the 15 guanine residues used to insert the cDNA into the Pst I site of pBR322. The 23-amino acid signal sequence is numbered -23 to -1 and the termination codon at nucleotide positions 782-784 is indicated by an asterisk. The canonical polyadenylation signal AATAAA located in the 3' untranslated region is underlined. The restriction enzyme sites that were 5' end labelled and used for DNA sequencing are shown. The nucleotide sequence of an extensive series of overlapping fragment was determined by the procedure of Maxam and Gilbert (18).

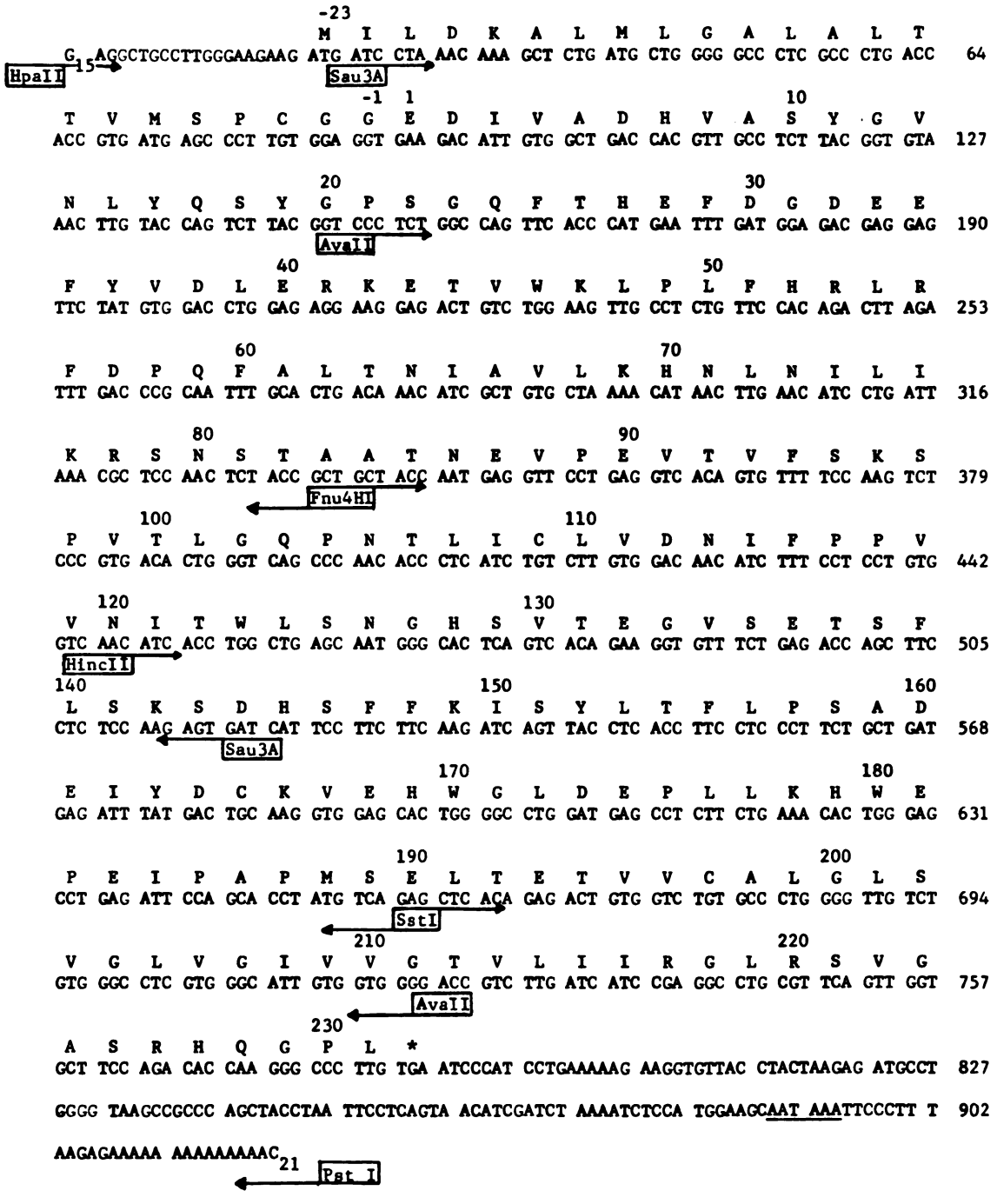
-23
M I L D K A L M L G A L A L T
G15 AGGCTGCCTTGGGAAGAAG ATG ATC CTA AAC AAA GCT CTG ATG CTG GGG GCC CTC GCC CTG ACC 64
HpaII
Sau3A
-1 1 10
T V M S P C G G E D I V A D H V A S Y G V
ACC GTG ATG AGC CCT TGT GGA GGT GAA GAC ATT GTG GCT GAC CAC GTT GCC TCT TAC GGT GTA 127
20 30
N L Y Q S Y G P S G Q F T H E F D G D E E
AAC TTG TAC CAG TCT TAC GGT CCC TCT GGC CAG TTC ACC CAT GAA TTT GAT GGA GAC GAG GAG 190
AvaII
40 50
F Y V D L E R K E T V W K L P L F H R L R
TTC TAT GTG GAC CTG GAG AGG AAG GAG ACT GTC TGG AAG TTG CCT CTG TTC CAC AGA CTT AGA 253
60 70
F D P Q F A L T N I A V L K H N L N I L I
TTT GAC CCG CAA TTT GCA CTG ACA AAC ATC GCT GTG CTA AAA CAT AAC TTG AAC ATC CTG ATT 316
80 90
K R S N S T A A T N E V P E V T V F S K S
AAA CGC TCC AAC TCT ACC GCT GCT ACC AAT GAG GTT CCT GAG GTC ACA GTG TTT TCC AAG TCT 379
Fnu4HI
100 110
P V T L G Q P N T L I C L V D N I F P P V
CCC GTG ACA CTG GGT CAG CCC AAC ACC CTC ATC TGT CTT GTG GAC AAC ATC TTT CCT CCT GTG 442
120 130
V N I T W L S N G H S V T E G V S E T S F
GTC AAC ATC ACC TGG CTG AGC AAT GGG CAC TCA GTC ACA GAA GGT GTT TCT GAG ACC AGC TTC 505
HincII
140 150 160
L S K S D H S F F K I S Y L T F L P S A D
CTC TCC AAG AGT GAT CAT TCC TTC TTC AAG ATC AGT TAC CTC ACC TTC CTC CCT TCT GCT GAT 568
Sau3A
170 180
E I Y D C K V E H W G L D E P L L K H W E
GAG ATT TAT GAC TGC AAG GTG GAG CAC TGG GGC CTG GAT GAG CCT CTT CTG AAA CAC TGG GAG 631
190 200
P E I P A P M S E L T E T V V C A L G L S
CCT GAG ATT CCA GCA CCT ATG TCA GAG CTC ACA GAG ACT GTG GTC TGT GCC CTG GGG TTG TCT 694
SstI
210 220
V G L V G I V V G T V L I I R G L R S V G
GTG GGC CTC GTG GGC ATT GTG GTG GGG ACC GTC TTG ATC ATC CGA GGC CTG CGT TCA GTT GGT 757
AvaII
230
A S R H Q G P L *
GCT TCC AGA CAC CAA GGG CCC TTG TGA ATCCCAT CCTGAAAAAG AAGGTGTTAC CTACTAAGAG ATGCCT 827
GGGG TAAGCCGCCC AGCTACCTAA TTCCTCAGTA ACATCGATCT AAAATCTCCA TGGAAGCAAT AAATTCCCTT T 902
AAGAGAAAAA AAAAAAAAAC21
Pst I

Figure 1

Fig. 2. Comparison of the protein sequences of DR and DSα subunits predicted from cDNA clones. The DR$_\alpha$ protein sequence was predicted from the nucleotide sequence of a DRα cDNA clone previously isolated (19), and the protein sequence of the DSα subunit was predicted from the nucleotide sequence of the pDS$_\alpha$-12 cDNA clone described here. The DRα and DSα subunits may be organized into a leader (L) region, two extracellular domains (α1 and α2), a connecting peptide (CP), a transmembrane (TM) region and an intracytoplasmic (CY) region based on the studies by Korman et al. (10), Auffray et al. (8) and Lee et al. (11). Homologous regions of DR and DSα subunits are boxed. The homology between these two molecules is 21.7% in the leader sequence, 48.2% in the α 1 domain, 65.2% in the α2 domain and 72.5% in the combined CP, TM and CY regions. The numbering system is based on the DS sequence.

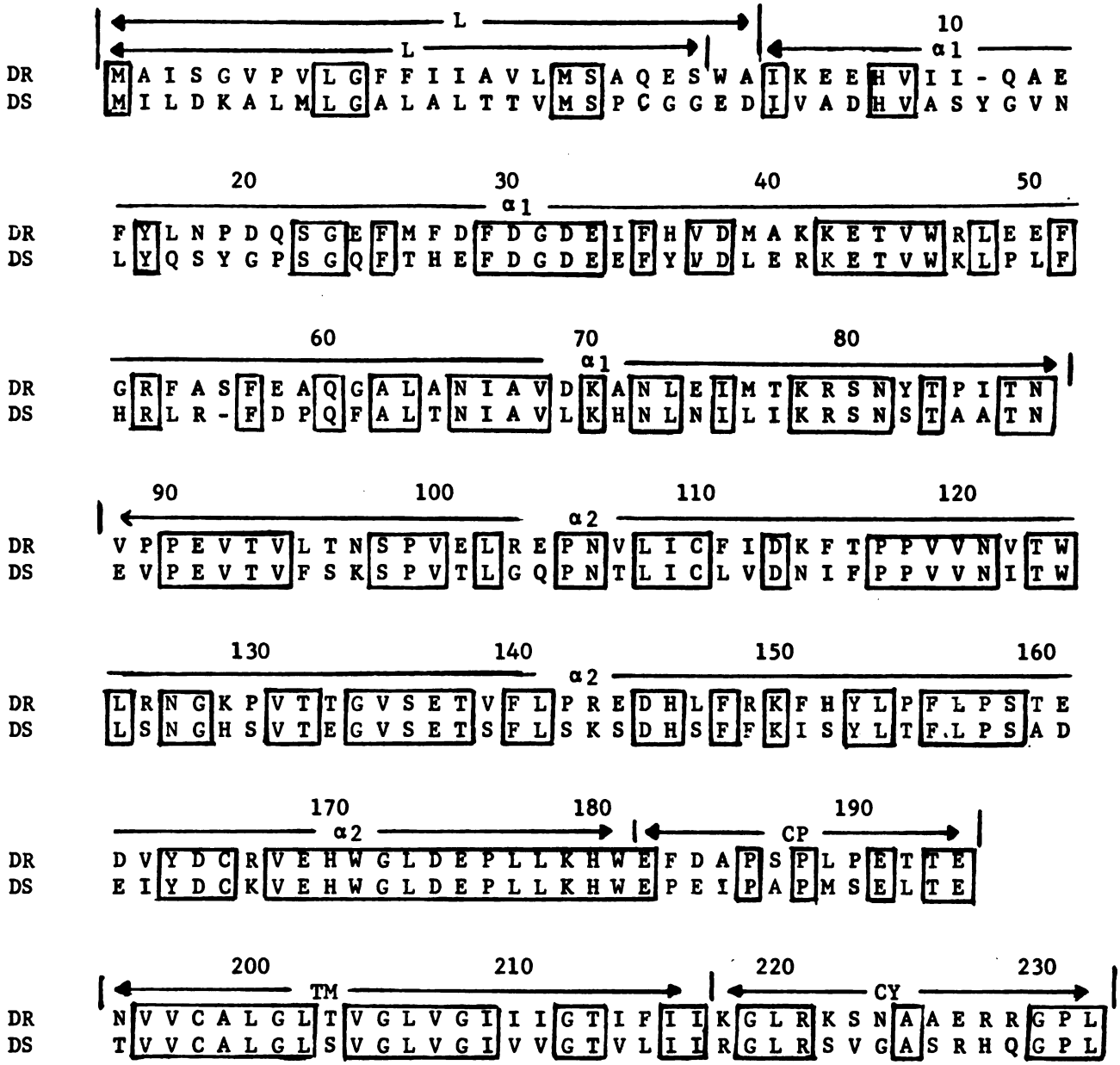
L
L
10
α1
DR M A I S G V P V L G F F I I A V L M S A Q E S W A I K E E H V I I - Q A E
DS M I L D K A L M L G A L A L T T V M S P C G G E D I V A D H V A S Y G V N
20
30
α1
40
50
DR F Y L N P D Q S G E F M F D F D G D E I F H V D M A K K E T V W R L E E F
DS L Y Q S Y G P S G Q F T H E F D G D E E F Y V D L E R K E T V W K L P L F
60
70
α1
80
DR G R F A S F E A Q G A L A N I A V D K A N L E I M T K R S N Y T P I T N
DS H R L R - F D P Q F A L T N I A V L K H N L N I L I K R S N S T A A T N
90
100
α2
110
120
DR V P P E V T V L T N S P V E L R E P N V L I C F I D K F T P P V V N V T W
DS E V P E V T V F S K S P V T L G Q P N T L I C L V D N I F P P V V N I T W
130
140
α2
150
160
DR L R N G K P V T T G V S E T V F L P R E D H L F R K F H Y L P F L P S T E
DS L S N G H S V T E G V S E T S F L S K S D H S F F K I S Y L T F L P S A D
170
α2
180
190
CP
DR D V Y D C R V E H W G L D E P L L K H W E F D A P S P L P E T T E
DS E I Y D C K V E H W G L D E P L L K H W E P E I P A P M S E L T E
200
TM
210
220
CY
230
DR N V V C A L G L T V G L V G I I I G T I F I I K G L R K S N A A E R R G P L
DS T V V C A L G L S V G L V G I V V G T V L I I R G L R S V G A S R H Q G P L

Figure 2

of 51 (24%) of the residues are identical and amino acid positions 47-59, where only 4 of 13 (31%) of the residues are identical. In addition, the 3' untranslated regions show no significant homology above random (30%).

When the nucleotide sequence of the pDS$_\alpha$-12 cDNA insert, which codes for the DS7$_\alpha$ subunit, is compared with that of the pDCH1 cDNA insert, which codes for a nearly complete DS4,6α subsunit, 11 nucleotide substitutions as well as deletion of a codon at nucleotide positions 254-256 are observed (Fig. 3). A comparison of the predicted protein sequences for the DSα and DS4,6α subunits (Fig. 4) reveals 11 amino acid sequence differences, 10 of which are located whithin the N-terminal portion of the molecule which by analogy to the DRα subunit has been designated the α1 domain (10,11). The α2 domain, which represents an immunoglobulin-like region conserved among class I and class II histocompatibility antigens (12-15), is identical for both the DS4, 6 and DS7 α subunits. The 11th amino acid sequence difference is found in a region of the molecule referred to as the connecting peptide and bridges the α2 domain and the transmembrane segment. It is intrguing that 4 of the amino acid differences are clustered at position 47-56, one of several regions, as pointed out above, where DR and DS subunits differ substantially. Thus, there is even a suggestion of a hypervariable region. Indeed, recent studies by M. Davis et al. (personal communication) reveal that the murine Aα subunits (homologues of DS) show a high degree of variability at these same positions. Note also that of nine nucleotide differences within the α1 domain (the deletion of the codon at positions 254-256 is considered as a single substitution event), seven result in amino acid differences. This high precentage of non-silent substitutions suggests that there are fewer structural and functional constraints on the α1 domain than for other

Fig. 3. Comparison of the DNA sequences of DS7 and DS4, 6 α subunits. The upper line represents the DS7α DNA sequence and lower line the DS4, 6 α DNA sequence (from ref. 8). The sequences are identical unless otherwise indicated. Positions 254-256 repesent a deletion in DS7 α with respect to DS4, 6α . The first N-terminal amino acid is indicated by an arrow and the termination codon by an asterisk.

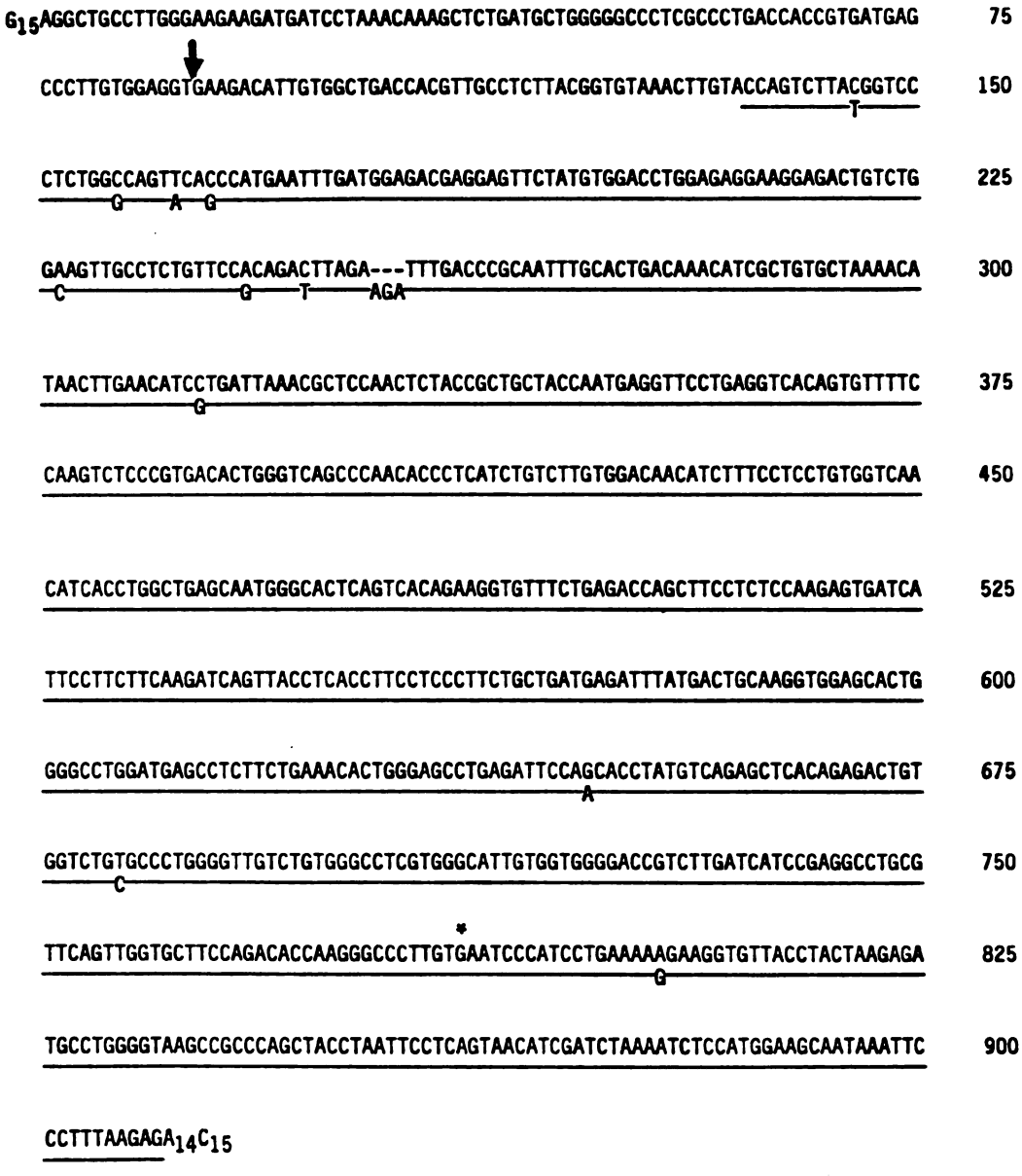
G15AGGCTGCCTTGGGAAGAAGATGATCCTAAACAAAGCTCTGATGCTGGGGGCCCTCGCCCTGACCACCGTGATGAG 75
CCCTTGTGGAGGTGAAGACATTGTGGCTGACCACGTTGCCTCTTACGGTGTAAACTTGTACCAGTCTTACGGTCC 150
T
CTCTGGCCAGTTCACCCATGAATTTGATGGAGACGAGGAGTTCTATGTGGACCTGGAGAGGAAGGAGACTGTCTG 225
G A G
GAAGTTGCCTCTGTTCCACAGACTTAGA---TTTGACCCGCAATTTGCACTGACAAACATCGCTGTGCTAAAACA 300
C G T AGA
TAACTTGAACATCCTGATTAAACGCTCCAACTCTACCGCTGCTACCAATGAGGTTCCTGAGGTCACAGTGTTTTC 375
G
CAAGTCTCCCGTGACACTGGGTCAGCCCAACACCCTCATCTGTCTTGTGGACAACATCTTTCCTCCTGTGGTCAA 450
CATCACCTGGCTGAGCAATGGGCACTCAGTCACAGAAGGTGTTTCTGAGACCAGCTTCCTCTCCAAGAGTGATCA 525
TTCCTTCTTCAAGATCAGTTACCTCACCTTCCTCCCTTCTGCTGATGAGATTTATGACTGCAAGGTGGAGCACTG 600
GGGCCTGGATGAGCCTCTTCTGAAACACTGGGAGCCTGAGATTCCAGCACCTATGTCAGAGCTCACAGAGACTGT 675
A
GGTCTGTGCCCTGGGGTTGTCTGTGGGCCTCGTGGGCATTGTGGTGGGGACCGTCTTGATCATCCGAGGCCTGCG 750
C
*
TTCAGTTGGTGCTTCCAGACACCAAGGGCCCTTGTGAATCCCATCCTGAAAAAGAAGGTGTTACCTACTAAGAGA 825
G
TGCCTGGGGTAAGCCGCCCAGCTACCTAATTCCTCAGTAACATCGATCTAAAATCTCCATGGAAGCAATAAATTC 900
CCTTTAAGAGA14C15

Figure 3

Fig. 4. Comparison of protein sequences of DS4, 6 and DS7 α subunits. The predicted protein sequence of the DS7α subunit (Fig. 1) is compared with the predicted protein sequence of the DS4, 6α subunit taken from ref. 8. As the DS4, 6 α cDNA clone did not contain the first 17 amino acids, residues 1-17 are based on protein sequence analysis (4). Domain organization is as in Fig. 2. There are 11 amino acid differences between these two molecules, 10 in the α1 domain, the 11th in the connecting peptide region.

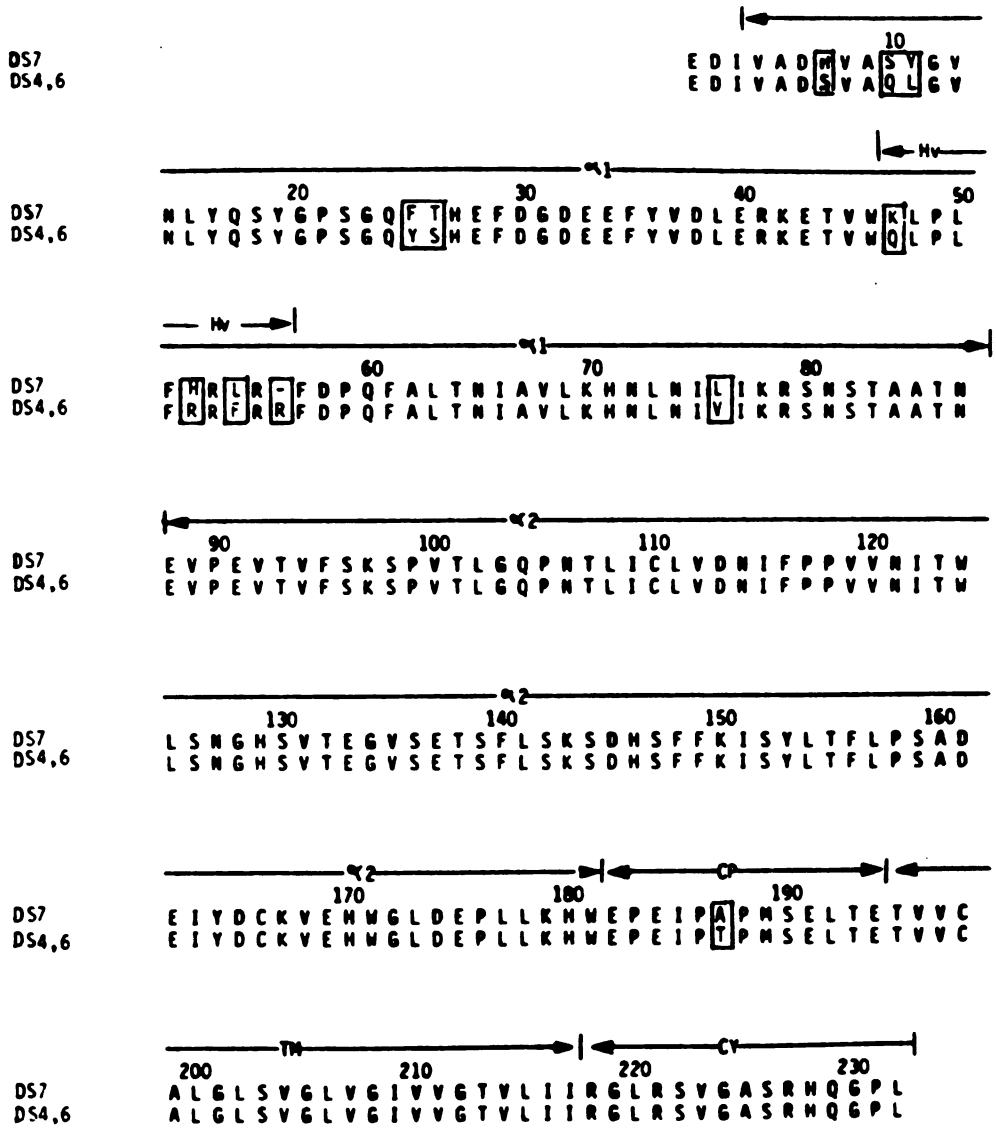
DS7
DS4,6
10
E D I V A D H V A S V G V
E D I V A D S V A Q L G V
Hv
α1
20 30 40 50
N L Y Q S Y G P S G Q F T H E F D G D E E F Y V D L E R K E T V W K L P L
N L Y Q S Y G P S G Q Y S H E F D G D E E F Y V D L E R K E T V W Q L P L
Hv
α1
60 70 80
F H R L R - F D P Q F A L T N I A V L K H N L N I L I K R S N S T A A T N
F R R F R R F D P Q F A L T N I A V L K H N L N I V I K R S N S T A A T N
α2
90 100 110 120
E V P E V T V F S K S P V T L G Q P N T L I C L V D N I F P P V V N I T W
E V P E V T V F S K S P V T L G Q P N T L I C L V D N I F P P V V N I T W
α2
130 140 150 160
L S N G H S V T E G V S E T S F L S K S D H S F F K I S Y L T F L P S A D
L S N G H S V T E G V S E T S F L S K S D H S F F K I S Y L T F L P S A D
α2
CP
170 180 190
E I Y D C K V E H W G L D E P L L K H W E P E I P A P M S E L T E T V V C
E I Y D C K V E H W G L D E P L L K H W E P E I P T P M S E L T E T V V C
TM
CY
200 210 220 230
A L G L S V G L V G I V V G T V L I I R G L R S V G A S R H Q G P L
A L G L S V G L V G I V V G T V L I I R G L R S V G A S R H Q G P L

Figure 4

proteins. We believe that pDSα-12 and pDCH1 represent products of alleles at the same locus because 126 of 127 nucleotides at the 3' untranslated (UT) regions are identical. Although, at this time we cannot entirely rule out the possibility that these clones are products of different loci that have recently diverged (16, 17), it is difficult to see how they could have accumulated so many differences in the $\alpha 1$ domain while maintaining a relatively constant 3' UT region. Additional studies should resolve this issue.

Thus, the DSα subunit consists of an N-terminal domain with a high degree of variability and a highly conserved immunoglobulin-like domain. The constancy of the $\alpha 2$ domain may reflect a functional requirment such as the ability to interact with the immunoglobulin-like domain of the DS β subunit, interactions analogous to those that occur between immunoglobulin heavy and light chains. The role of variability in the α 1 domain, however, remains to be determined.

REFERENCES

1. Allison J.P., L.E. Walker, W.A. Russell, M.A. Pellegrino, S. Ferrone, R.A. Reisfeld, J.A. Frelinger, and J. Silver. Proc. Natl. Acad. Sci. USA 75, 3953-9356 (1978).
2. Goyert, S.M. and J. Silver. Nature (London) 294, 266-268 (1981).
3. Goyert, S.M., J.E. Shively and J. Silver. J. Exp. Med. 156, 550-566 (1982).
4. Bono, R. and J.L. Strominger. Nature (London) 299, 836-838 (1982).
5. Silver, J. and S. Ferrone. Nature (London) 279, 436-437 (1979).
6. Silver, J. and S. Ferrone. Immunogenetics 10, 295-298 (1980).
7. Goyert, S.M., J.J. Hubert, R.A. Curry and J. Silver. J. Hum. Immun. 1, 161-175 (1980).
8. Auffray, C., A.J. Korman, M. Roux-Dosseto, R. Bono and J.L. Strominger. Proc. Natl. Acad. Sci. USA 79. 6337-6341 (1982).
9. Moriuchi, T., H.C. Chang, R. Denome and J. Silver. Nature (London) 301, 80-82 (1983).
10. Korman, A.J., C. Auffray, A. Schamboeck and J.L. Strominger. Proc. Natl. Acad. Sci. USA 79, 6013-6017 (1982).
11. Lee, J.S., J. Trowsdale, P.L. Travers, J. Carey, F. Grosveld, J. Jenkins and W.F. Bodmer. Nature (London) 99, 750-753 (1982).
12. Orr, H.T., J.A. Lopez de Castro, D. Lancet and J.L. Strominger. Biochemistry 18, 5711-5720 (1979)
13. Malissen, M., B. Malissen and B.R. Jordan. Proc. Natl. Acad. Sci. USA 79, 893-897 (1982)

14. Larhammer, D., K. Gustafsson, L. Claesson, P. Bill, K. Wiman, L. Schenning, J. Sundelin, E. Widmark, P. Peterson and L. Rask. Cell 30, 153-161 (1982).

15. Larhammer, D., L. Schenning, K. Gustafsson, K. Wiman, L. Claesson, L. Rask and P.A. Peterson. Proc. Nalt. Acad. Sci. USA 79, 3687-3691 (1982).

16. Trowsdale, J., J. Lee, F. Grosveld, J. Bodmer and W. Bodmer. Proc. Natl. Acad. Sci. USA 80, 1977-1976 (1982)

17. Auffray, C. EMBO J. 2, 121-124 (1983)

18. Maxam, A.M. and W. Gilbert. Meth. Enzym. 65, 499-560 (1980)

19. Kajimura, Y., H. Toyoda, J.E. Shively, S. Goyert, J. Silver and K. Itakura. DNA (submmitted).

Chapter III

THY-1 cDNA sequence suggests a novel regulatory mechanism

By

Tetsuya Moriuchi, Hsiu-Ching Chang,

Roger Denome and Jack Silver

Department of Microbiology and Public Health

Michigan State University

E. Lansing, MI 48824

Reprinted by permission from Nature, 301, 80-82

Thy-1 was originally defined in mice as a cell-surface alloantigen of thymus and brain with two allelic forms, Thy-1.1 and Thy-1.2 (ref. 1). Subsequently, the Thy-1.1 alloantigenic determinant was identified in rats (2). In both species, Thy-1 is present in large amounts on thymus and brain cells (3) and in smaller quantities on the fibroblasts (4), epidermal cells (5), mammary glands (6) and immature skeletal muscle (7). In many of these tissues the level of Thy-1 expression changes dramatically during cell differentiation. The molecules expressing the Thy-1 antigenic determinant have been isolated from rat and mouse brain cells and have been shown to have a molecular weight of 17,500 (ref. 8). One-third of the Thy-1 molecule is carbohydrate and the remainder is a polypeptide of 111 amino acids whose sequence has been fully determined (9). We report here the isolation and characterization of a cDNA clone encoding the rat thymus Thy-1 antigen but find that the DNA sequence ends prematurely at a position corresponding to amino acid 103. It appears to be a complete transcipt, however, as the last codon is followed directly by a poly(A) tract.

Poly(A) containing RNA was isolated from W/Fu rat thymocytes, and cDNA was synthesized by reverse trancription using oligo(dT)12-18 primer. Double-stranded cDNA was synthesized as described previously (10). A cloned cDNA library was constructed by inserting the total cDNA population into the Pst I site of pBR322 using poly(dG).poly(dC) homopolymeric extensions (11). Colonies containing cDNA were screened (12) with a synthetic ^{32}P-labelled oligodeoxynucleotide (17-mer) mixture composed of all 32 possible sequence permutations corresponding to amino acids 82-87 of Thy-1. Two hydridizaton-positive clones were isolated from 10,000 colonies. One clone (pT64) contained cDNA encoding the Thy-1 antigen and

its entire sequence was determined by procedure of Maxam and Gilbert (13). The overall structure of the cDNA clone pT64 and the sequencing strategy used are shown in Fig. 1.

The DNA sequence contains a 45-base 5'-untranslated region followed by 57 bases coding for a presumptive leader peptide of 19 amino acids starting at the first methionine codon (numbered -19 in Fig. 2). The leader sequence exhibits several features characteristic of leader peptide present at the amino terminus of membrane bound proteins (14); it contains many hydrophobic amino acids (11 non-polar residues) and terminates in a residue with a small neutral side chain (glycine). The remainder of the predicted amino acid sequence is in agreement with the published amino acid sequence for rat brain Thy-1 (9). There are, however, several unusual features at the 3' end of the Thy-1 cDNA clone. The DNA sequence ends prematurely at a position corresponding to amino acid 103, 8 amino acids earlier than predicted from the protein sequence, but followed directly by a poly(A) tract. There is, however, no termination codon. Furthermore, a presumptive polyadenylation signal, 5'-AATAAA-3'(ref. 15), which is part of the coding sequence, is found 12 nucleotides upstream from the poly(A) tract. Two alternative explanations for these unusual features at the 3' end can be proposed: (1) a deletion of the sequence encoding the C-terminal peptide occurred during the cloning procedure while retaining the sequence corrsponding to the poly(A) tract, or (2) the AATAAA sequence at positions 293-298 was recognized as a polyadenylation signal <u>in vivo</u> and position 310, which is 12 nucleotides downstream from the AATAAA sequence, served as a polyadenylation site.

The first explanation cannot be excluded although such a cloning artefact has not been described before. The second explanation is a very

Fig. 1. Partial restriction map and strategy for sequencing Thy-1 cDNA clone, pT64. The entire cDNA was cleaved at selected sites with restriction endonuclease and fragments corresponding to the insert were purified by acrylamide gel electrophoresis and electroelution. Fragments were treated with calf intestinal or bacterial alkaline phosphatse and ^{32}P-labelled at both 5' ends with T4 polynucleotide kinase as described (13). Alternatively, some fragment were ^{32}P-labelled at both 3' ends with DNA polymerase (26). Double-end-labelled fragments were separated by acrylamide gel electrophoresis, electro-eluted and subjected to partial chemical degradation sequence analysis as described by Maxam and Gilbert (13). Partially cleaved fragments were separated on 8, 15 and 20% acrylamide gels. The extent of sequence determined from each fragment is indicated by the length of the arrow. The closed circles and vertical lines at one end of each arrow indicate labelling at the 3' end or 5' end, respectively. The sequencing strategy used here allowed us to sequence both DNA strands completely.

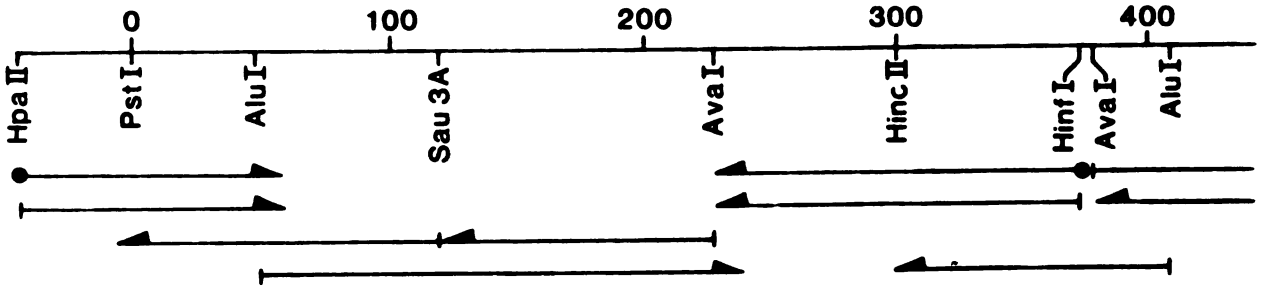
0
100
200
300
400
Hpa II
Pst I
Alu I
Sau 3A
Ava I
Hinc II
Hinf I
Ava I
Alu I

Figure 1

Fig. 2. Primary structure and predicted amino acid sequence for pT64. Amino acids are numbered relative to the amino-terminus of the protein sequencec as detertmined by Campbell et al (9). The 19-amino acid leader sequence 5' to the codon for the first amino acid is numbered -19 to -1. The hexanucleotide sequence, AATAAA, at the 3' end is underlined. A 24-base poly(A) tail directly follows codon 103.

Primary structure and predicted amino acid sequence for pT64

```
                                                        -19
                                                        met asn pro val
CTGCAAGCTAGGGGAGCCCAGACCCAGGACGGAGCTATTGGCACC  ATG AAC CCA GTC

                                                                -1  1
ile ser ile thr leu leu leu ser val leu gln met ser arg gly gln
ATC AGC ATC ACT CTC CTG CTT TCA GTC TTG CAC ATG TCC CGA GGA CAG

                                    10
arg val ile ser leu thr ala cys leu val asn gln asn leu arg leu
AGG GTG ATC AGC CTG ACA GCC TGC CTG GTG AAC CAG AAC CTT CGA CTG

         20                                         30
asp cys arg his glu asn asn thr asn leu pro ile gln his glu phe
GAC TGC CGT CAT GAG AAT AAC ACC AAC TTG CCC ATC CAG CAT GAG TTC

                         40
ser leu thr arg glu lys lys lys his val leu ser gly thr leu gly
AGC CTG ACC CGA GAG AAG AAG AAG CAC GTG CTG TCG GGC ACC CTG GGG

 50                                      60
val pro glu his thr tyr arg ser srg val asn leu phe ser asp arg
GTT CCC GAG CAC ACT TAC CGC TCC CGC GTC AAC CTT TTC ACT GAC CGG

                  70                                        80
phe ile lys val leu thr leu ala asn phe thr thr lys asp glu gly
TTT ATC AAG GTC CTT ACT CTA GCC AAC TTC ACC ACC AAG GAT GAG GGC

                                    90
asp tyr met cys glu leu arg val ser gly gln asn pro thr ser ser
GAC TAC ATG TGT GAA CTT CGA GTC TCG GGC CAG AAT CCC ACA AGC TCC

        100         103
asn lys thr ile asn val
AAT AAA ACT ATC AAT GGT (A)24
```

Figure 2

attractive model for a mechanism regulating rapid changes of Thy-1 gene expression during cell differentiation and is similar to that observed in other systems. For example, during differentiation, B lymphocytes undergo a shift from expression of membrane-bound IgM to secreted IgM. It has been shown that the transcription unit for secreted and membrane-bound IgM heavy chains contains two separate poly(A) addition sites. The basis for selection between these two gene products apparently resides in the choice of polyadenlyation site of the precursor RNA followed by differential "splicing out" of introns (16, 17). Expression of the late adenovirus transcription unit is alo controlled through differential poly(A) site selection (18). A similar mechanism may be responsible for regulating Thy-1 gene expression. Differential poly(A) site selection may control the expression of the C-terminal coding segment, which is apparently the region of Thy-1 responsible for membrane integration (9) and, in the presence of differentiation-inducing factors such as thymic hormones, complete expression of cell-surface Thy-1 may be induced. Komuro and Boyse demonstrated that Thy-1 negative precursor cells in the bone marrow and spleens of mice can be induced to differentiate <u>in vitro</u> into Thy-1 positive T lymphocytes within 2 h of incubation with crude thymus extract (16). Our hypothesis might explain the mechanism responsible for this rapid induction of Thy-1 on the cell surface.

DNA and amino acid sequence analyses have revealed homology among the major histocompatibility antigens, β2-microglobulin and immunoglobulins, and suggests that all three have evolved from a common ancestral gene encoding a primitive domain (20-22). Homology between Thy-1, which has a domain-like structure including a disulphide loop of the appropriate size, and immunoglobulins has been noted at the amino acid sequence level (23).

We have compared the nucleotide sequences of Thy-1 cDNA amd mouse λ1 light chain variable region (24) which was shown to have the highest degree of amino acid sequence homology to Thy-1 (23). When the sequences are aligned as shown in Fig. 3, overall homology between the neculeotide sequence of Thy-1 and the variable region is 36%. This sequence homology supports the hypothesis that Thy-1 and immunoglobulin have evolved from a common ancestral gene. Furthermore, our observation concerning the Thy-1 cDNA clone, pT64, suggests that the Thy-1 gene may even have inherited a remnant of a mechanism involved in regulating immunoglobulin gene expression, namely differential polyadenylation. Additional experiments aimed at defining the genetic organization and regulating elements of Thy-1 expression are in progress. Thus Thy-1 which has long been used primarily as a marker for T lymhpocyte differentiation (25), may itself represent an intriguing system for study of gene expression.

Fig. 3. Comparison of the DNA sequences of mouse $V\lambda 1$ (MOPC104E) and rat Thy-1. DNA sequences were aligned with the amino acid squence alignment of Williams et al. (23), except for certain regions, indicated by parentheses, that were realigned to increase homology. Identical amino acid residues are underlined and nucleotide sequence homologies are indicated by vertical lines. Gaps inserted to maximize homology are indicated by dashes. The hypervariable regions (HV1, HV2 and HV3) are lined above the amino acid sequence and the Variable-Constant (V-C) junction (J region) is indicated by a dotted line. It is interesting that although the overall nucleotide sequence homology between Thy-1 and $V\lambda 1$ is 36%, comparisons of small segments reveal considerable variation in the degree of homology. The homology is 41% in positions 1-20, 19% in the position 21-34 (sequences corresponding to the first hypervariable region 1), 41% in positions 35-85, 38% in positions 86-92 (HV3) and 24% in the positions 93-103 (sequences corresponding to the J region).

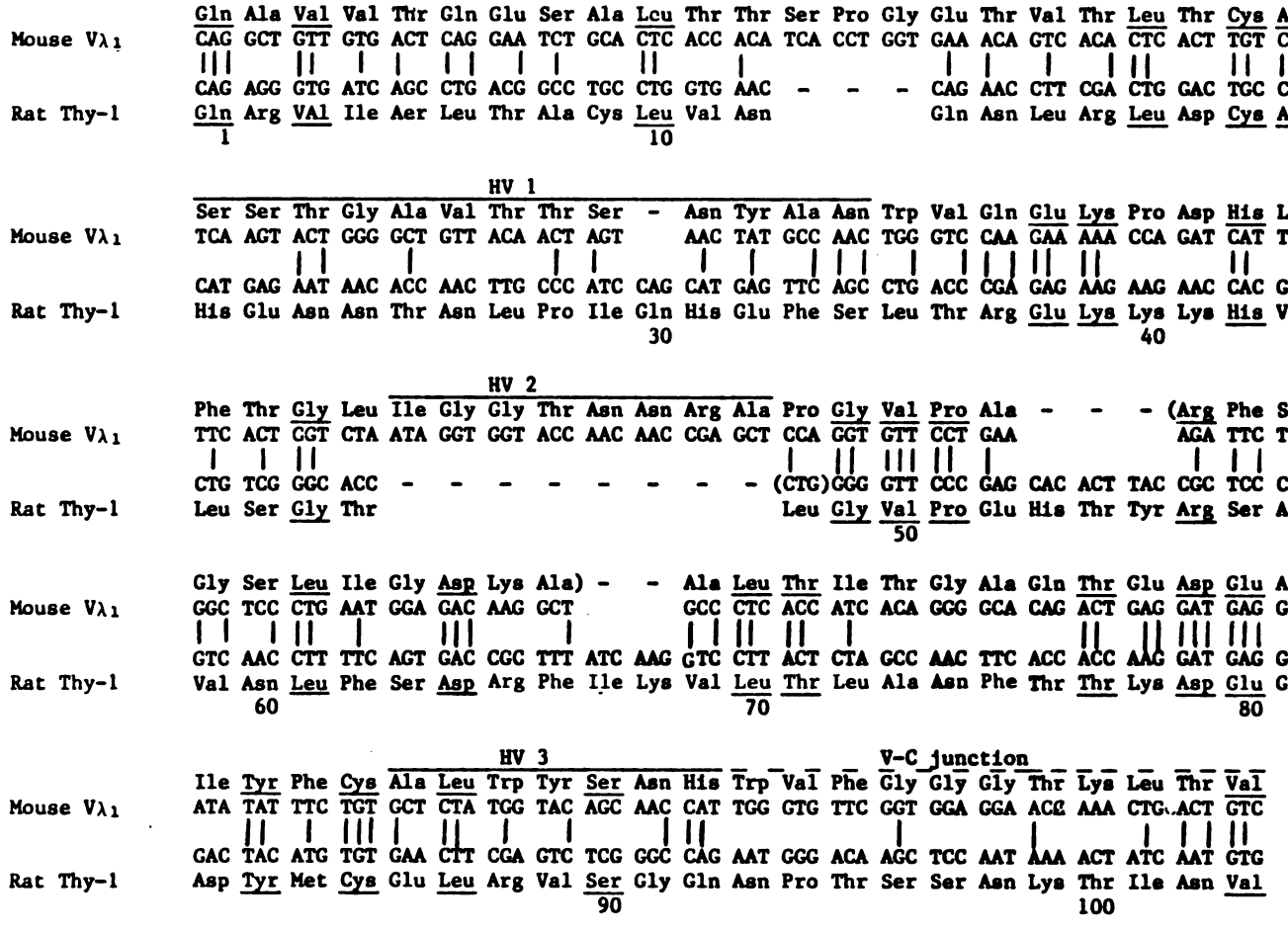

Figure 3

REFERENCES

1. Reif, A.E. and Allen, J.M.V. J. Exp. Med. 120, 413-433 (1964).
2. Douglas T.C. J. Exp. Med. 136, 1054-1062 (1972).
3. Reif, A.E. and Allen, J.M.V. Nature 209, 523 (1966).
4. Stern. P.I. Nature 246, 76-78 (1973).
5. Scheid, M., Boyse, E.A., Carswell, E.A. and Old, L.J. J. Exp. Med. 135, 938-955 (1972).
6. Lennon, V.A., Unger, M. and Dulbecco, R. Proc. Natl. Acad. Sci. USA 75, 6093-6097 (1978).
7. Lesley, J.F. and Lennon, V.A. Nature 268, 163-165 (1977).
8. Kuchel, P.W., Campbell, D.G., Barclay, A.N. and Williams, A.F. Biochem. J. 169, 411-417 (1978).
9. Campbell, D.G., Gagnon, J., Reid, B.M. and Williams, A.F. Biochem. J. 195, 15-30 (1981).
10. Efstradiadis, A., Kafatos, F.C., Maxam, A.M. and Maniatis, T. Cell 7. 279-288 (1976).
11. Villa-Komaroff, I., Estratiadis, A., Broome, S., Lomedico, P., Tizard, R., Naber, S.P., Chick, W.L. and Gilbert, W. Proc. Natl. Acad. Sci. USA 75, 3727-3731 (1978).
12. Suggs, S.V., Wallace, R.B., Hirose, T., Kawashima, E.H. and Itakura, K. Proc. Natl. Acad. Sci. USA 78, 6613-6617 (1981).
13. Maxam, A.M. and Gilbert, W. Meth. Enzym. 65, 499-560 (1980).
14. Inouye, M. and Halegoua, S. CRC Crit. Rev. Biochem. 7, 339-371 (1980).

15. Proudfoot, N.J. and Brownless G.G. Nature 263, 211-214 (1976).

16. Rogers, J., Early, P., Cater, C., Calame, K., Bond, M., Hood, L. and Wall, R. Cell 20, 303-312 (1980).

17. Early, P., Rogers, J., Davis, M., Clame, K., Bond, M., Wall, R. and Hood, L. Cell 20, 313-319 (1980).

18. Nevins, J.R. and Wilson, N.C. Nature 290, 113-118 (1981).

19. Komuro, K. and Boyse, E.A. Lancet i, 740-743 (1973),

20. Peterson, P.A., Cunningham, B.A., Berggard, I. and Edelman, G.M. Proc. Natl. Acad. Sci. USA 69, 1697-1701 (1972).

21. Steinmetz, M, Frelinger, J., Fisher, D., Hunkapiller, T. Pereira, D., Weissman, S.M., Uehara, H. Nathenson, S. and Hood, L. Cell 24, 125-134 (1979).

22. Larhammer, D.L., Schenning, L., Gustafsson, K., Wiman, K., Claesson, L., Rask, L. and Peterson, P.A. Proc. Natl. Acad. Sci. USA 79, 3687-3691 (1982).

23. Williams, A.F. and Gagnon, J. Science 216, 696-703 (1982).

24. Bernard, O., Hozumi, N. and Tonegawa, S. Cell 15, 1133-1144 (1978).

25. Raff, M.C. Immunology 19, 637-650 (1970).

26. Maniatis, T., Fritsch, E.F. and Sambrook, R. Molecular Cloning, 115 (Cold Spring Harbor Loboratory, New York, 1982).

Chapter IV

THE NUCLEOTIDE SEQUENCE OF THE MOUSE THY-1 GENE

Hsiu-Ching Chang and Jack Silver

Department of Microbiology and Public Health

Michigan State University

E. Lansing, MI 48824

Running Title: Genomic cloning and DNA sequence analysis of the mouse Thy-1.2 gene

SUMMARY

The mouse Thy-1.2 gene was isolated from a C57BL/6 cosmid library. The nucleotide sequence of the mouse Thy-1.2 gene has been determined from an 8-Kb long Eco RI fragment. The predicted amino acid sequence indicates that mouse Thy-1.2 antigen contains a 19 amino acid leader peptide and the 112 amino acids previously reported by protein sequence analysis, plus 31 extra amino acids at the carboxyl terminus. These 31 amino acids contain a stretch of 20 amino acids, at positions 124-143, which is highly hydrophobic. Northern blot analysis of RNA from mouse tissues indicates that the sequence coding for these 31 amino acids is present on poly(A) containing RNA of brain and thymus tissues. The entire coding sequence is distributed among three exons, encoding for amino acid residues -19 to -8, -7 to 106 and 107 to 143, respectively. The predicted amino acid sequences and gene organizations of mouse, rat and human Thy-1 were compared. The sequences are more highly conserved between rodents than between rodents and human. The carboxyl termini are highly conserved among these three species, with homologies of more than 85 %. The highly conserved nature of the carboxyl terminus suggests a functional role for this fragment of the molecule.

INTRODUCTION

The Thy-1 antigen was originally defined in mice as a cell surface alloantigen of thymus and brain with two allelic forms, Thy-1.1 and Thy-1.2 (Reif and Allen, 1964). Subsequently, The Thy-1.1 determinant was identified in rat (Douglas, 1972). In both species, Thy-1 is present in large amounts on the surface of thymus and brain cells, with about 600,000 molecules per cell on rat thymocytes, and equivalent amounts on brain cells . Thus Thy-1 is probably the most abundant surface glycoprotein of both cell types (Acton et al., 1974). In rodents, Thy-1 is also expressed in small amounts on fibroblasts (Stern, 1973), epidermal cells (Scheid et al., 1972), mammary glands (Lennon, 1978) and immature skeletal muscle cells (Lesley and Lennon, 1977). In many of these tissues the level of Thy-1 expression changes dramatically during cell differentiation.

The molecules expressing the Thy-1 antigenic determinant have been isolated from rat and mouse brain cells (Letarte-Muirhead et al., 1975, Trowbridge et al., 1975). Thy-1 has been shown to have a molecular weight of 17,500, of which one third of the molecule is carbohydrate, and the remainder is a polypeptide of 111 amino acids in rat (Campbell et al., 1981) and 112 amino acids in mouse (Williams and Gagnon, 1982). Glycoproteins that are structurally related to rat and mouse brain Thy-1 have been purified from the brain of human (Cotmore et al, 1981), dog (Mckenzie and Fabre, 1981), chicken (Rostas et al., 1983) and frog (Mansour and Cooper, 1984), and even from invertebrate squid (Williams and Gagnon, 1982). The expression of Thy-1 seems to be conserved in the

evolution of neuroral cells and fibroblasts; however, the expression of Thy-1 on lymphoid cells varies in different species (Dalchou and Fabre, 1979). Thus, although Thy-1 is the most abundant surface molecule on rat and mouse thmocytes, the human homolog cannot be detected unambiguously on human thymocytes. Furthermore, Thy-1 is present on mouse T lymphocytes but absent from B cells. In constrast, Thy-1 is absent from rat T cells, but present on rat bone marrow cells (Hunt et al., 1978).

The Thy-1 molecule was previously thought to have an unusual mode of integration within the membrane. It can be heavily labelled in the thymocyte membrane with an affinity label which labels only the hydrophobic section of the membrane proteins (Owen et al., 1980). Previous biochemical studies had failed to reveal any extended sequence of hydrophobic amino acid that might function as a transmembrane segment. Furthermore, the C-terminal peptide had hydrophobic properties despite a lack of a hydrophobic sequence (Kuchel et al., 1978). Therefore, it was suggested that the hydrophobic properties of the C-terminal peptide were due to the linkage of lipid which anchored Thy-1 on the cell membrane. However, recent studies of Thy-1 cDNA clones by T. Moriuchi (personal communication) have shown that the rat Thy-1 molecule is 142 amino acids long with a hydrophobic segment at the carboxyl terminus. This suggests that Thy-1 in fact is an integral protein anchored on the membrane via a hydrophobic segement.

Thy-1 has long been used as a mouse T cell marker. Despite much speculation, however, there is no clear understanding of its biological significance. On the basis of amino acid sequence, secondary and tertiary structures (Cohen et al., 1978) and nucleotide sequence (Moriuchi et al., 1983), Thy-1 shows striking structural homologies with Ig domains, and

with other members of the Ig superfamily (Larhammer et al., 1982, Korman et al., 1982). This has led to the hypothesis that Thy-1 is likely to be a primordial domain from which all the members of the Ig superfamily might have evolved. It would therefore be interesting to analyze the evolution of the Thy-1 gene in relation to other members of Ig superfamily.

In this study, we report the isolation of mouse Thy-1 genomic clones from a C57BL/6 comsid library, and DNA sequence analysis of the Thy-1 gene. The predicted amino acid sequences and gene organization of mouse, rat and human genes are compared. The protein sequences and gene organization of the Thy-1 genes are well conserved among the three species, with homologies in the third exon of greater than 85%.

RESULTS

Isolation and charavterization of mouse Thy-1 genomic clones.

A cosmid library of mouse C57BL/6 DNA was constructed in the c2RB vector as described (Bates and Swift, 1983) with a small modification. The C57BL/6 DNA was partially digested with Sau 3A to an average size of 40 Kb, and dephosphorylated before ligation with Bam HI-Sma I digested c2RB DNA. Ligated DNA was then packaged *in vitro*, and infected *E. coli* strain 1046 was used to construct a cosmid library (Figure 1). To isolate a mouse Thy-1 genomic clone, the cosmid library was probed with the P-labelled insert of the rat thymocyte Thy-1 cDNA clone, pT64 (Moriuchi et al., 1983). After screening 175,000 clones, three independent clones were isolated, cT10, cT34 and cT5F. The partial restriction enzyme maps of these clones are presented in Figure 2A. The cT10 clone shares only 4 and 2 Kb of 3' end of the Thy-1 gene with cT34 and cT5F, respectively. The total insert of these clones spans about 75 Kb DNA along the mouse DNA.

The clones were first identified as Thy-1 containing clones by Southern blot analyses, since the mouse Thy-1 gene had been shown to be a single copy gene, and to be located in an 8 kb Eco RI fragment in Southern blot analysis of total cellular DNA (data not shown). The Southern blot analysis of these cosmid clones shows that each contains only one hybridizing band. The cT10 clone contains an 8-kb Eco RI band of the exact size as indicated by the cellular DNA analysis. The Eco RI fragments of clones cT34 and cT10 were subcloned into the Eco RI site of pBR322. The Thy-1 containing subclones, pcT34 and pcT108, were subjected

Figure 1. Scheme for the use of c2RB vector in the construction of a mouse C57BL/6 cosmid library. After digestion with Bam HI and Sma I, c2RB DNA is ligated to phosphatase-treated mouse C57BL/6 DNA insert, which was partially digested with Sau 3A to an average size of approximately 40 Kb. Note that Bam HI and Sau 3A have the same cohesive GATC ends. A high ATP concentration is used during the ligation reaction to prevent the joining of the Sma I blunt ends. *In vitro* packaged DNA was used to infect *E. coli* 1046 in order to construct the mouse cosmid library.

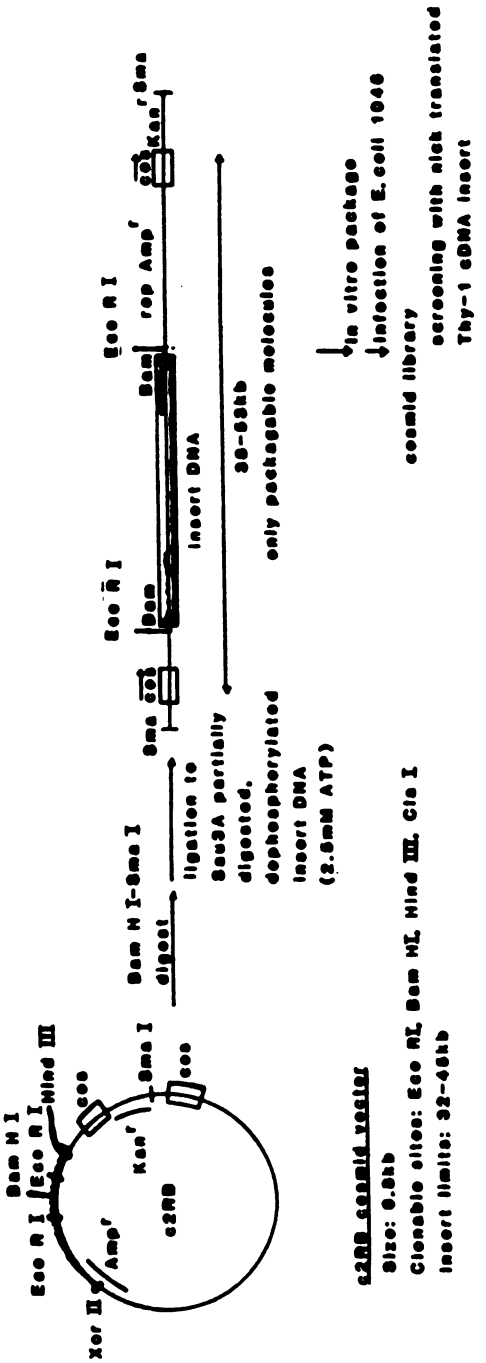
Bam H I
Eco R I
Eco R I
Hind III
Xor II
Amp'
Kan'
Sma I
ori
c2RB
Bam H I-Sma I
digest
ligation to
Sau3A partially
digested,
dephosphorylated
insert DNA
(2.5mM ATP)
Sma ori
Eco R I
Bam
insert DNA
Eco R I
Bam
rop Amp'
ori
Kan' Sma
30-45kb
only packagable molecules
in vitro package
infection of E.coli 1046
cosmid library
screening with nick translated
Thy-1 cDNA insert
c2RB cosmid vector
Size: 6.8kb
Cloning sites: Eco RI, Bam HI, Hind III, Cla I
Insert limits: 32-46kb

Figure 1

Figure 2. Restriction enzyme maps of Thy-1 genomic clones.
(A) Partial Eco RI digestion maps of Thy-1 containing clones, cT10, cT34 and cT5F are shown. Clone cT10 shares only 4 and 2 kb of the 3' end of the Thy-1 gene with cT34 and cT5F, respectively, making the total insert of these clones span 75 Kb of DNA along the mouse genome.
(B) Partial restriction map of the Thy-1 containing genomic clone, pcT108, and the intron-exon organization of the mouse Thy-1 gene. The thicker line, and the boxed area denote the coding region of Thy-1. The amino acids were numbered relative to the amino-terminus of the protein sequence as determined.
(C) The strategy for sequencing the Thy-1 gene. The pBR322 clones containing the Eco RI insert of Thy-1 were cleaved at selected sites with restriction endonucleases as indicated. The fragments were isolated and 5' end-labelled with T4 polynucleotide kinase, or 3' end-labelled by filling in with DNA polymerase I. Double-end labelled fragments were digested with restriction enzymes, separated on the agarose gel, eluted, and then subjected to DNA sequencing by Maxam and Gilbert's method. The extent of sequence determined from each fragment is indicated by the length of the arrow. The solid circle and vertical lines at one end of each arrow indicated labelling at the 3' or 5' end, respectively. The DNA sequence was completely determined by sequencing either both complementary strands or the same strand at least twice, starting with different cleavage sites.

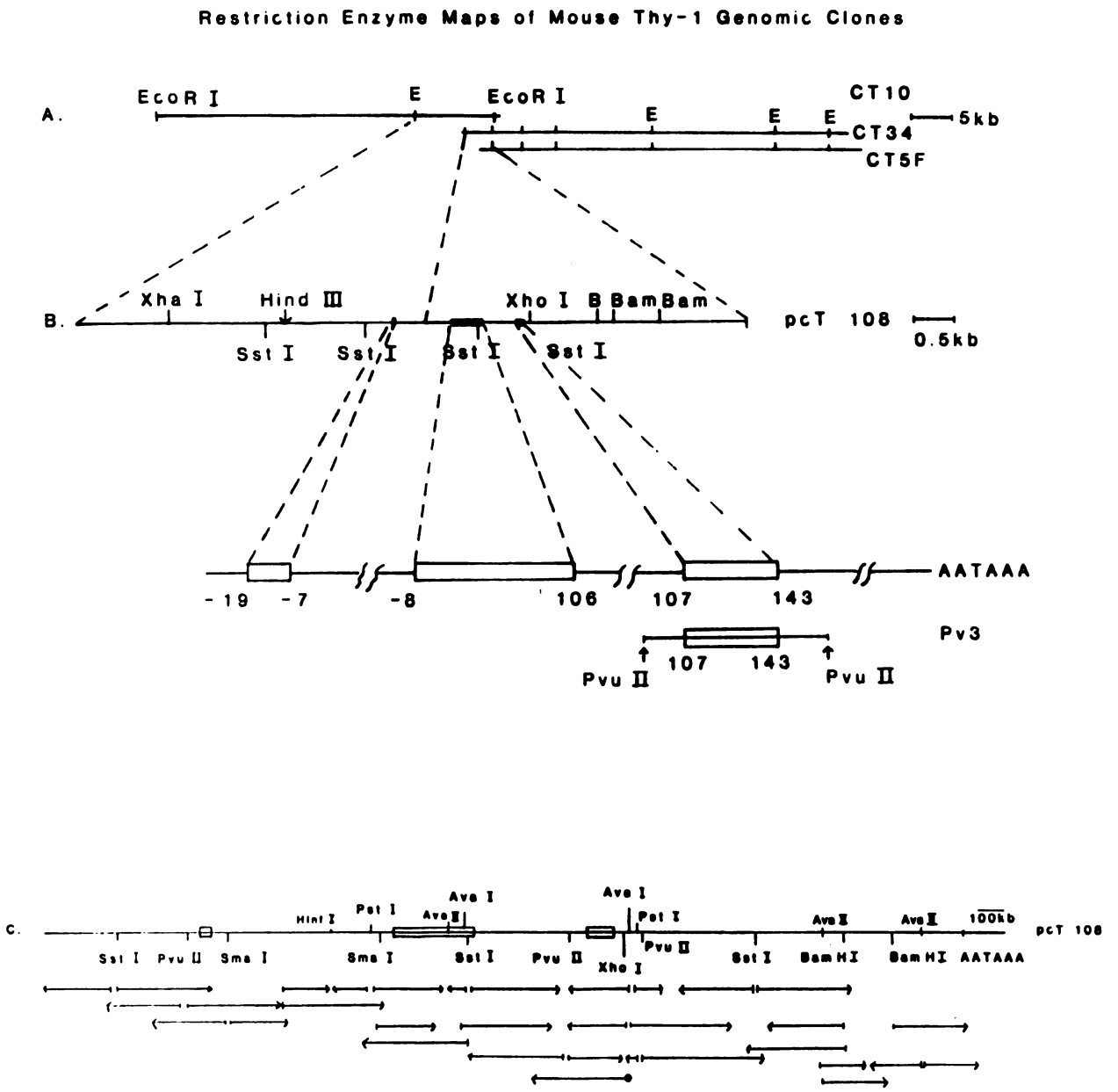
Restriction Enzyme Maps of Mouse Thy-1 Genomic Clones
A.
EcoR I
E
EcoR I
CT10
5kb
CT34
CT5F
B.
Xha I
Hind III
Xho I
B
Bam
Bam
pcT 108
0.5kb
Sst I
Sst I
Sst I
Sst I
AATAAA
-19
-7
-8
106
107
143
Pv3
Pvu II
107
143
Pvu II
C.
Hint I
Pst I
Ave I
Ave I
Ave I
Pst I
Ave I
Ave I
100kb
pcT 108
Sst I
Pvu II
Sma I
Sma I
Sst I
Pvu II
Pvu II
Xho I
Sst I
BamHI
BamHI
AATAAA

Figure 2

to DNA sequence analysis following the strategies illustrated in Figure 2C. The sequence data enabled us to definitively identify these clones as Thy-1 containing clones.

Structure of the mouse Thy-1 gene.

Comparison of the nucleotide sequences of mouse genomic and rat cDNA clones, and comparison of the predicted amino acid sequence with the protein sequence previously reported (Williams and Gagnon, 1982) established the intron-exon organization of the mouse Thy-1 gene (Figure 2B). The coding region of Thy-1 is divided into 3 exons by two introns of sizes 590 bp and 388 bp, respectively. As is often the case, the first exon codes for part of the 5' untranslated region of the mRNA and the first 12 amino acids of the signal peptide, amino acids -19 to -8, followed by a 590 bp intron. The second exon contains most of the amino acids of the mature protein, amino acids -7 to 106, followed by a 388 bp intron. Both introns were occurred by RNA spliced between the first and second base of the junctional codons according to GT/AG rule (Nevins, 1983). The rest of the predicted amino acid residues, positions 107-143, followed by the termination codon, TGA, and the 3' untranslated region of mRNA reside in the third exon. Since the only difference between the two mouse allelic forms, Thy-1.1 and Thy-1.2, is amino acid residue 89, in which Thy-1.1 is arginine and Thy-1.2 is glutamine, the gene we have cloned is the mouse Thy-1.2 gene. This is consistent with the serological typing of the C57BL/6 Thy-1 gene. The predicted protein sequence shows that the mouse Thy-1.2 gene contain a 19 amino acid presumptive leader peptide, 112 amino acid residues which fit completely with the protein sequence reported by Campbell et al. 1981, plus 31 additional amino acids

Figure 3. Sequence of the mouse Thy-1.2 gene. The nucleotide sequence and predicted protein sequence (three letter amino acid code) of the Thy-1 gene are shown. Exons are underlined, with a thin line for the untranslated region of mRNA, and with a thick line for the protein-coding sequence. The initiation codon is boxed and the termination codon is indicated by an asterisk. The AATAAA sequences are boxed.

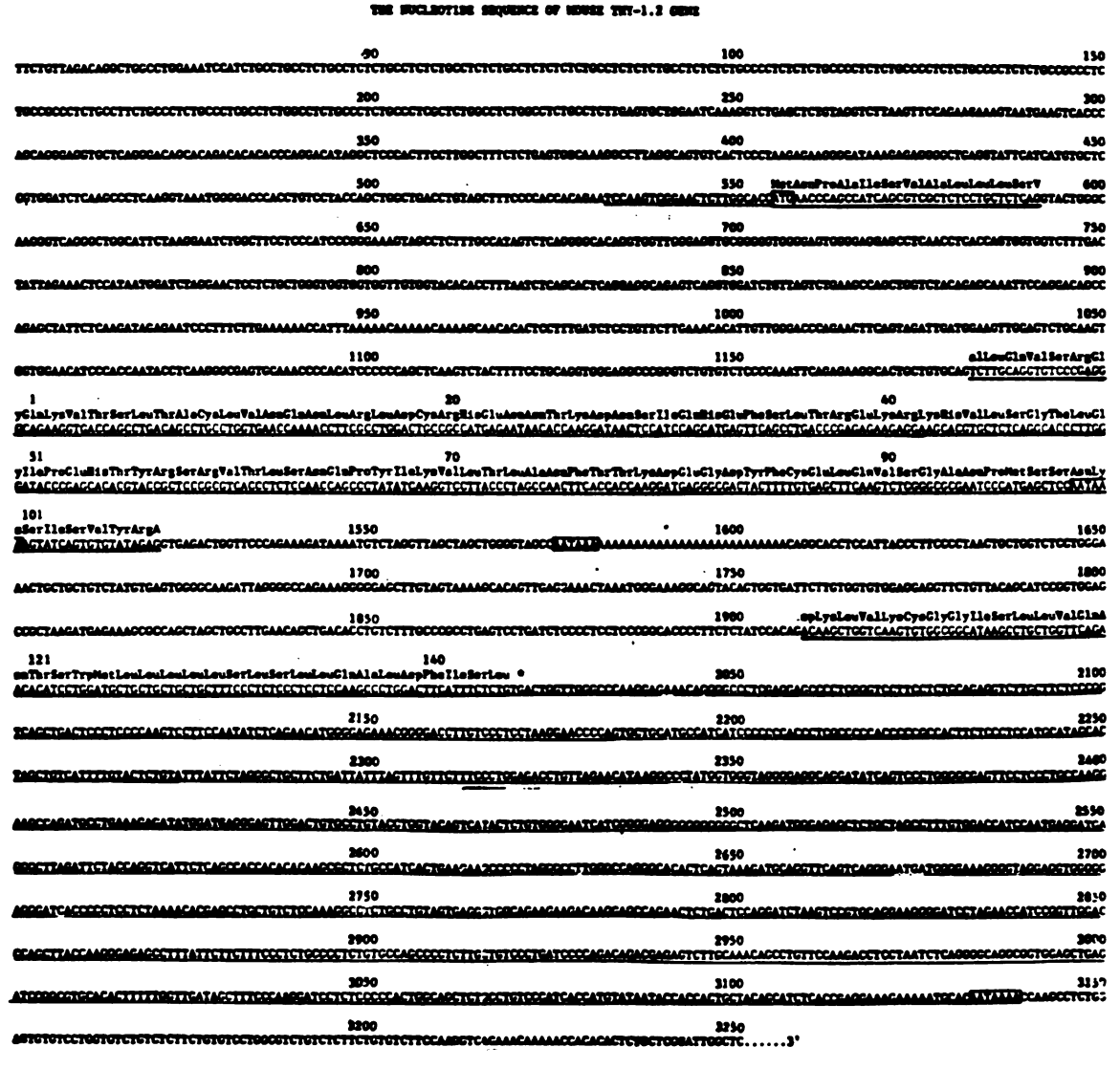

Figure 3

(Figure 3). The last 20 amino acids at the carboxyl terminus (positions 124 to 143) are highly hydrophobic and include ten leucine residues.

The gene organization of mouse, rat and human Thy-1.

Since the nucleotide sequences of rat and human Thy-1 genes have been determined (Seki et al., 1984), we were able to compare the gene organization of mouse, rat and human Thy-1 genes (Figure 4). The distribution of the coding sequence among the three exons is identical for all three species, coding for amino acids, -19 to -8, -7 to 106 and 107 to 143. All the introns follow RNA splicing occurs between the first and second base of the junctional codons accroding to GT/AG rule, with variations in length among these three species. This indicates that the gene organization of Thy-1 is highly conserved among these three species, with very well conserved exon structure, and slightly less conserved intron structures. However, there are some features conserved in the rat and human thy-1 genes, but not in the mouse Thy-1 gene. Namely, there is a deletion of one amino acid at position 29, and there are two AATAAA sequences in the 3' untranslated region of both human and rat Thy-1 gene located around 550 bp and 1100 bp downstream of the termination codon, TGA. However, there is only one AATAAA sequence in the 3' untranslated region of mouse Thy-1 gene, located at a position corresponding to the second AATAAA sequence of the rat and human gene.

Northern blot analysis of RNA from mouse and rat tissues

To determine whether the 31 extra amino acids predicted from the cDNA and gene structure are present in the mRNA, northern blot analyses were

Figure 4. Comparison of the gene organization of mouse, rat and human Thy-1. The intron-exon organization of rat and human Thy-1 genes were determined by Seki et al., 1984. The sizes of exons are shown by the amino acid residues, and the sizes of introns and 3' untranslated region are shown by the nucleotide base pairs. The locations of the AATAAA sequences are shown in non-proportional scale.

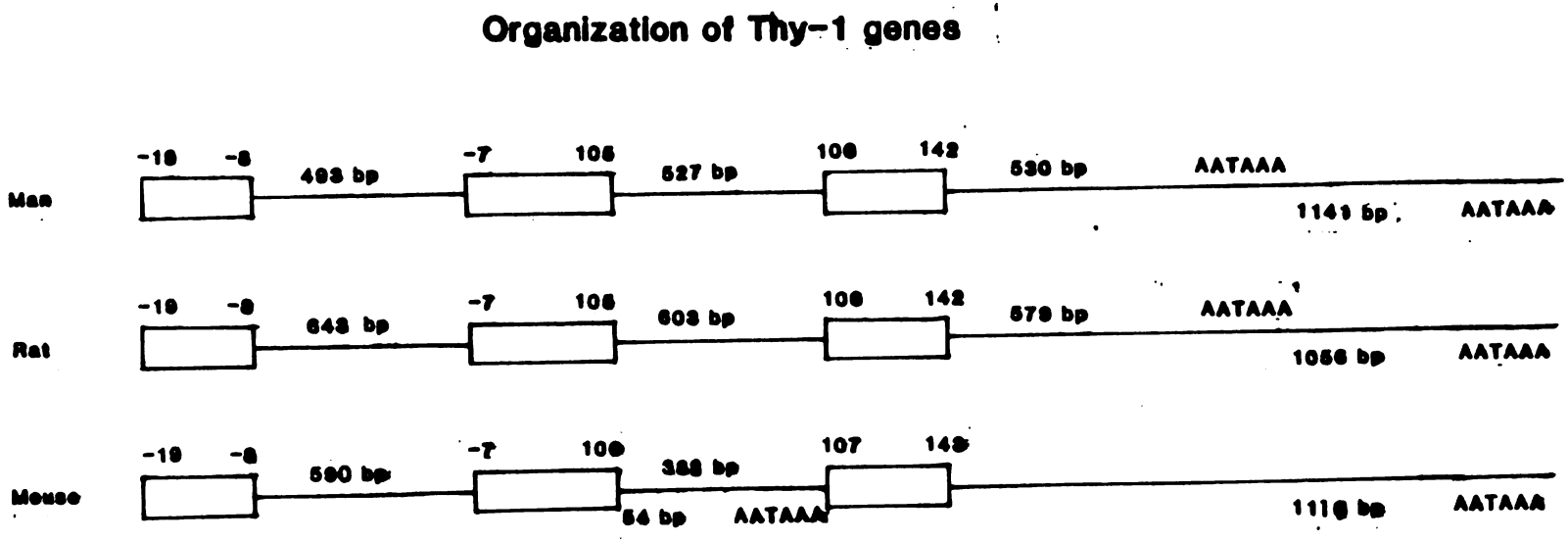
Organization of Thy-1 genes
Man
-19
-8
493 bp
-7
105
527 bp
106
142
530 bp
AATAAA
1141 bp
AATAAA
Rat
-19
-8
648 bp
-7
105
603 bp
106
142
579 bp
AATAAA
1056 bp
AATAAA
Mouse
-19
-8
590 bp
-7
106
388 bp
107
143
54 bp
AATAAA
1116 bp
AATAAA

Figure 4

performed. A northern blot of poly(A) containing RNA from mouse brain, liver, spleen and thymus, was hybridized with a Pvu II digested fragment, PV3, which codes for amino acid residues 107 to 143 (Figure 2B). Only one species of mRNA was detected in brain and thymus tissues (Figure 5A). This demonstrates that the 31 extra amino acids are transcribed. The same blot was then hybridized with the entire insert of pcT108 two months later. The same bands were detected again (data not shown). The northern blot of total RNA from rat thymus and form several human B cell lines was also probed with the insert of pT64. There is also only one species of Thy-1 mRNA in the rat thymus, with an approximate size 1.7 Kb.

Comparison of 5' untranslated region

To determine where the Thy-1 gene begins, the nucleotide sequences of the 5' ends of the rat cDNA and rat and mouse genomic clones were compared (Figure 6). The nucleotide sequences of the first coding exon and 28 base pairs upstream from the initiation codon, ATG, are identical for the rat cDNA and genomic DNA; upstream from this point, the sequences differ substantially. The same situation can be found when the nucleotide sequences of rat cDNA and mouse genomic DNA are compared. However, the homology between the sequences of the rat and mouse genomic clones extend far beyond this point. The rat gemonic nucleotide sequence CTGCAG/A at positions -30 to -24 fits very well with the consensus sequence of the splicing acceptor site, TNCAG/G (Nevins, 1983).This indicates the presence of an intron within the 5' untranslated region of both the rat and mouse Thy-1 genes. In the mouse thy-1 gene, a sequence close to the consensus sequence of the splicing acceptor site, at positions -30 to -25, can also be found. When a Southern blot of Eco RI-Hind III digested pcT108 was

Figure 5. Nothern blot analyses of RNA from mouse and rat tissues. (A) Poly(A) containing RNA was isolated from mouse brain (B), liver (L), spleen (S) and thymus (T), and subjected to northern blot analysis as described in Experimental Procedures.
(B) Total RNA of rat thymus (5) and several human B cell lines, 3105 (1), 3106(2), DR2-Ma (3), and DR2-Mo (4) was isolated, and analyzed as described above, except it was hybridized with the insert of pT64.

NORTHERN BLOT ANALYSIS OF RNA FROM VARIOUS TISSUES

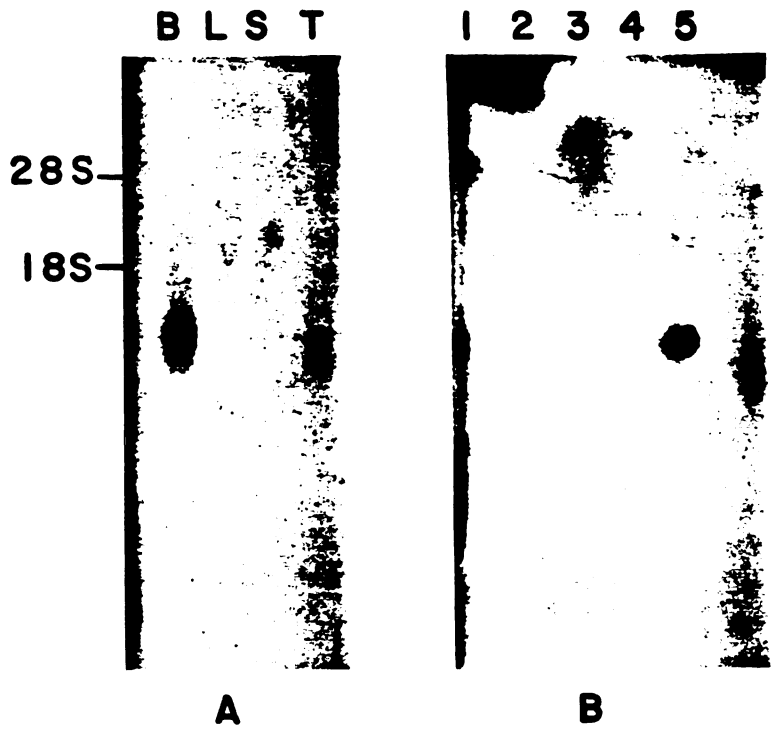

Figure 5

Figure 6. Comparison of the 5' end nucleotide sequences of rat cDNA and mouse and rat genomic DNA. The sequence of the rat cDNA (top line), starting from position -45 upstream from the initiation codon, ATG, is compared to the sequences of rat, and mouse genomic clones (middle line and bottom line, respectively). Identical bases are shown by a line between the sequences. The initiation codon is boxed, and the putative splicing points are marked.

COMPARISON OF 5'END NUCLEOTIDE SEQUENCES OF RAT AND MOUSE THY-1 GENOMIC DNA AND RAT cDNA

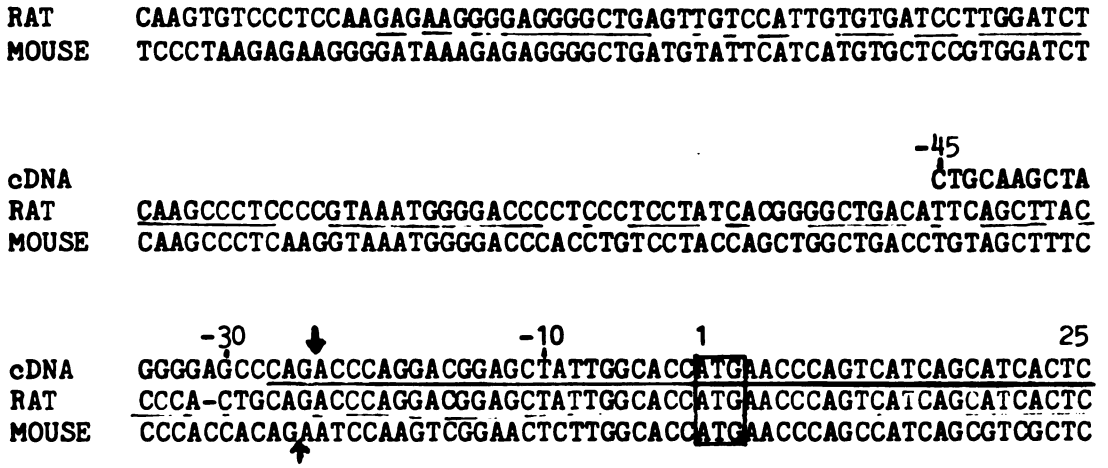

Figure 6

hybridized with a full length mouse Thy-1 cDNA clone (a gift form M. Davis), both fragments hybridized (data not shown). Since the only Hind III site in pcT108 is located 1.5 Kb upstream from the initiation codon, it suggests that there is an intron in the 5' untranslated leader of the Thy-1 gene with an estimated size of 1.5 kb. The lack of restriction enzyme sites within this region, suggesting the presence of highly repetitive sequences, has made this region difficult to sequence. Introns in the 5' untranslated region have been observed in the genes of rat insulin I (Cordell et al., 1979) and chicken ovalbumin (Breathenach et al., 1978) with sizes of approximately 120 bp. The reason for the existence of such a large intron in the 5' untranslated region of the Thy-1 gene is unknown. Whether it plays some role in regulating the expression of Thy-1 remains to be determined.

Comparison of the predicted amino acid sequences of the Thy-1 genes

The predictd protein sequences of mouse, rat and human Thy-1 are aligned with a gap inserted at amino acid 29 in both rat and human Thy-1 to maximize the homology (Figure 7). All three Thy-1 antigens contain a 19 amino acid leader peptide plus the extra 31 amino acids at the carboxyl end previously undetected by other investigators. Although, there is a high degree of overall homology, there are three regions, located at positions 23 to 29, 60 to 68, and 99 to 105, that vary substantially among the species. It is intriguing to note that these regions correspond closely to the locations of the hypervariable regions of Ig V domains (Bernard et al., 1978), suggesting that as in immunoglobulins there is little structural constraint in these regions of the molecule.

The degree of protein sequence homology between the three species are

Figure 7. Comparison of the predicted amino acid sequences of mouse, rat and human Thy-1. Identical amino acid residues are underlined. The hydrophobic segment is underlined.

Protein sequence comparison of mouse, rat, and human Thy-1

```
        -19                                  -10
Mouse   met asn pro ala ile ser val ala leu leu leu ser val leu gln
Rat     -----------val---------ile-thr--------------------------
Human   --------leu-ala---------ile-ala-------------thr-----lys-----

                -1   1                                        10
Mouse   val ser arg gly gln lys val thr ser leu thr ala cys leu val
Rat     met----------------arg-----ile--------------------------
Human   val----------------lys-----thr--------------------------

                                     20
Mouse   asn gln asn leu arg leu asp cys arg his glu asn asn thr lys
Rat     ------------------------------------------------------asn
Human   asp-----ser-----------------------------------thr-ser-ser

                    30                                       40
Mouse   asp asn ser ile gln his glu phe ser leu thr arg glu lys arg
Rat     leu-pro     ----------------------------------------lys
Human   ser-asn     --------tyr-----------------------------thr-lys

                                     50
Mouse   lys his val leu ser gly thr leu gly ile pro glu his thr tyr
Rat     ----------------------------------val-------------------
Human   ----------------phe---------val-----val-------------------

                    60                                       70
Mouse   arg ser arg val thr leu ser asn gln pro tyr ile lys val leu
Rat     ----------------asn-----phe-ser-asp-arg-phe---------------
Haman   ------------thr-asn-phe-phe-ser-lys-tyr-asn-met-----------

                                     80
Mouse   thr leu ala asn phe thr thr lys asp glu gly asp tyr phe cys
Rat     ------------------------------------------------------met----
Human   tyr-----ser-ala---------ser-----------------thr-----thr----

                    90                                      100
Mouse   glu leu gln val ser gly ala asn pro met ser ser asn lys ser
Rat     --------arg-------------gln---------thr-----------------thr
Human   ala-----his-his---------his-ser-----pro-ile-----ser-gln-asn

                                    110
Mouse   ile ser val tyr arg asp lys leu val lys cys gly gly ile ser
Rat     ----asn-----ile-------------------------------------------
Human   val-thr-----ile-----------------------------glu-----------

                    120                                     130
Mouse   leu leu val gln asn thr ser trp met leu leu leu leu leu ser
Rat     ---------------------------------leu-----------------------
Human   --------ala-------------------leu-----------------------

                                    140
Mouse   leu ser leu leu gln ala leu asp phe ile ser leu
Rat     --------phe-------------thr-------------------
Human   --------leu-------------leu---------met--------
```

Figure 7

Degree of Thy-1 protein sequence homology

	1st exon	2nd exon	3rd exon	hydrophobic tail
rat/mouse	75%	81%	92%	85%
rat/human	66%	66%	89%	90%
mouse/human	75%	60%	86%	85%

Table 1. Degree of predicted Thy-1 protein sequence homology. The degree of homology between the predicted protein sequences of rat/mouse, rat/human and mouse/human are compared between individual exons.

compared exon by exon in Table 1. The overall protein sequence homology between mouse and rat Thy-1 is 80%, but only 66% between mouse and human and rat and human. Within the first exon, amino acid resiudes -19 to -8, the homology is 75%, 66% and 75%; thus, as expected the signal peptide is not conserved between the species. The homology for the second exon coding for amino acids -8 to 106, is 80% between mouse and rat, and 66% and 60% between both rodents and man, consistent with the evolutionary relationship of the species. The homologies within the third exon, encoding amino acids 107 to 143, are 92%, 89% and 86%. If only the hydrophobic segments are compared, the homologies are greater than 85% among all three species. Strong homologies in the hydrophobic regions of other molecules including membrane IgG1, IgG2a, and IgM (Tyler et al., 1982), and glycophorins of man, ox, and pig (Murayama et al., 1982) have been reported. This suggests that the structure of the hydrophobic segment is conserved in order to serve an important and specific, biologic function.

DISCUSSION

Analysis of the mouse Thy-1 genomic clone, and comparison of the protein sequences derived from the nucleotide sequences of mouse, rat and human Thy-1 genes have revealed several interesting features: the mouse Thy-1 antigen contains 143 amino acids instead of the 112 originally reported. The coding region of the Thy-1 gene is distributed into three exons, with the third exon encoding amino acids 107 to 143. The gene contains several AATAAA sequences through out the entire Thy-1 sequence, but only one, located 1110 bp downstream of the termination codon, TGA, serves as a polyadenylation signal. There is an intron in the 5' untranslated leader with a size of at least 1.5 kb. The overall structure of the mouse, rat and human Thy-1 genes is well conserved, including the intron-exon organization, the protein sequence and the 31 additional amino acids at the carboxyl terminus. There are three regions where protein sequence differences are concentrated when Thy-1 molecule from different species are compared, and these regions correspond closely to those of the hypervariable regions of the Ig V domain.

It should be noted that there are several inconsistencies between the DNA and protein sequences of Thy-1 (Table 2). For instance, the predicted amino acid compositions of mouse, rat and human Thy-1 (Table 2, line b.) contain tryptophan, but protein sequences of mouse, rat and human brain Thy-1 antigens, or even the squid brain glycoprotein (Table 2, line a.) do not. Also amino acid composition analyses failed to reveal any hydrophobic amino acids in the C-terminal peptide (Campbell et al., 1981;

Williams & Gagnon, 1982; Cotmore et al., 1981). These C-terminal hydrophobic regions are highly homologous (greater than 85%) among all three species, and are also present on the mRNA of rat and mouse thymocyte and brain tissues. According to the length and hydrophobicity of this segment, this region may serve as a transmembrane segment anchoring Thy-1 on the cell surface (Chou and Fasman, 1978). It is likely therefore that Thy-1 is in fact an integral membrane protein anchored on the membrane via a hydrophobic C-terminal segment.

In addition to Thy-1, there are two other proteins which are reported to have discrepancies between the biochemically determined protein sequences and the protein sequences predicted from the nucleotide sequence. One of those is a membrane protein, the Variant Surface Glycoproteins (VSGs) of the parasitic protozan trypanosome (Boothroyd et al., 1982). The nucleotide sequence from the cDNA clones of VSGs suggests that the primary translation product of one VSG gene contains a 23 amino acids hydrophobic tail at the carboxyl terminus which is not found on the isolated mature glycoproteins. Another one is a secretory protein, rat α-lactoalbumin (α-LA), which has been found to have 17 extra hydrophobic amino acids beyond the C-terminus (Dandekar and Qasba, 1981). Two forms of rat α-LA have been found with different molecular weights; however, only one species of mRNA can be found.

Several possible explanations have been proposed for these discrepancies. It is possible that the extra hydrophobic segments are removed during the processing of protein in a manner similar to the cleavage of the N-terminal signal peptide (Lingappa et al., 1978). An alternative explanation is that proteolytic degradation has occured during the preparation and purification of molecules and their peptide products,

Comparison of amino acid compositions of the Thy-1 antigens

	Human Brain Thy-1		Rat Brain Thy-1		Mouse Brain Thy-1.2		Squid Brain Glycoprotein
	a	b	a	b	a	b	a
CYS	5	4	4	4	4	4	8
ASX	9	10	15	17	14	15	15
THR	11	13	10	12	10	11	7
SER	12	19	8	13	11	16	7
GLX	11	12	10	12	11	12	12
PRO	5	3	3	3	3	3	3
GLY	5	5	4	6	4	6	8
ALA	4	5	2	3	3	4	5
VAL	8	9	9	10	8	9	8
MET	1	2	1	1	1	2	2
ILE	2	4	5	7	4	6	5
LEU	11	23	13	24	12	24	6
TRP	0	1	0	1	0	1	0
TYR	5	5	2	2	4	4	4
PHE	4	6	4	6	3	4	4
HIS	5	6	4	4	4	4	4
LYS	8	6	8	8	9	9	9
ARG	6	7	9	9	7	9	6

Line a. is the amino acid composition obtained from protein analysis (residue/molar amount).
Line b. is the amino acid composition predicted from nucleotide sequence (residue/molar amount).
Data were taken from references: human Thy-1 (Cotmore at al., 1981); rat Thy-1 (Campbell at al., 1981); mouse Thy-1.2 and squid brain glycoprotein (Williams and Gagnon, 1982); predicted amino acid compositions of human and rat Thy-1 (Seki et al., 1984).

or that the C-terminal peptide was lost simply because of its hydrophobicity. Further protein sequence studies and searches for possible precursors should resolve this issue.

EXPERIMENTAL PROCEDURES

DNA cloning

The cosmid library of mouse C57BL/6 was constructed in a cosmid vector c2RB as described (Bates and Swift, 1983) with a minor modification. The mouse C57BL/6 DNA was partially digested with Sau 3A to an average size of about 40 Kb, then dephosphorylated before ligated into the Bam HI and Sma I digested c2RB DNA. A high ATP concentraction is used during the ligation reaction to prevent the joining of the Sma I blunt ends. In vitro packaged DNA was used to infect E. coli strain 1046 to construct mouse libraries. The mouse library was screened using the insert of the Thy-1 cDNA clone as a hybridization probe. The DNA of putative Thy-1 positive clones was purified by mini preparation. The DNAs were cleaved with Eco RI, and other appropriate restriction enzymes and run on a 0.6% agarose gel for Southern blot analyses. Confirmed Thy-1 positive cosmid clones were then digested with Eco RI and subcloned into the Eco RI site of pBR322. The Subcloned DNA was used to transfom E. coli strain HB101, and the colonies were screened with the original Thy-1 probe. The general procedures for the growth, screening and analysis of cosmid libraries, nick translation, and Southern blot analyses were performed as described (Maniatis et al., 1982).

DNA preparation.

The preparation of DNAs from putative Thy-1 clones was perfomed using the alkaline-extract method (Birnboim and Doly, 1979) with some modifications.

A bacterial culture (40 ml) was cultured overnight without any amplification. The bacteria were pelleted, and treated with the GTC solution (50 mM glucose, 10 mM EDTA, 25 mM Tris-HCl, pH 8.0) without lysozyme. A clear lysate was prepared by the addition of 2 volumes of 0.2 M NaOH-1% SDS, and the cellular DNA was precipitated by the addition of 0.5 volume of 3 M potassium acetate (pH 5.2). The plasmid DNA remaining in the supernant was precipitated with 0.6 volume of isopropanol. The DNA was suspended in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0), treated with DNase-free RNase, phenol extracted, and then precipitated with 2.5 volume of ethanol. The DNA isolated from this procedure can be used for endonuclease restriction enzyme digestion, and ligation for subcloning. Large scale DNA preparation was perfomed as described above, except that CsCl gradent centrifugation was used for purification of the plasmid prior to RNase treatment and phenol extraction.

DNA sequence analysis

The Thy-1 positive subclones, pcT108 and pcT34, were subjected to nucleotide sequence analyses. The entire genomic clone was cleaved at selected sites with the restriction endonucleases, and fragments corresponding to the insert were purified by agarose gel electrophoresis and electroelution, or electrophoresis on low melting point agarose (BRL) and extraction with 0.3 M NaOAC (pH 5.2) saturated phenol. Fragments were dephosphorylated with calf intestinal alkaline phosphatase and end-labelled with T4 polynucleotide kinase. Only one of the fragments was end labelled by filling in with the DNA polymerase I. Double-end labelled fragments were digested with restriction enzymes, separated on agarose gel or low melting point agarose, and eluted from gels, then subjected to

partial chemical degradation sequence analysis as described (Maxam and Gilbert, 1980).

RNA analysis

The preparation of total RNA from mouse and rat tissues was perfomed as previously described (Chirgwin et al, 1979). The poly(A) containing RNA from mouse tissuses was isolated as described (Aviv and Leder, 1972). The northern blot analysis was performed as described (Lehrach et al, 1977). The mouse poly(A) containing RNA was run on a 1.2% formaldehyde agarose gel, botted onto a nitrocellulose filter, and hybridized with ^{32}P-labelled, PV3, in 50% formamide solution at 42°C. The northern blot analysis of RNA from rat thymus and several human B cell lines was done in the same way as that of mouse, except that it was hybridized with the ^{32}P-labelled insert of pT64 in a solution of 6x SSC, 10x Denhardt's, 0.1% SDS at 68°C.

ACKNOWLEDGEMENTS

The authors thank Drs. Jerry Dodgson and Paul Bates for their advice in cloning technology, Drs. Tetsuya Moriuchi, Tetsunori Seki and Roger Denome for their kindly providing their unpublished data, and Dr. Susan Cornad for her helpful discussion.

REFERENCES

Acton, R. T., Morris, R. J. and Williams, A. F. (1974) Estimation of the amount and tissue distribution of rat Thy-1.1 antigen. Eur. J. Immunol. 4, 598-602.

Aviv, H. and Leder, P. (1972) Purification of biologically active globin messenger RNA by chromotography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. U.S.A. 69, 1408-1413.

Bates, P. L. and Swift, R. A. (1983). Double cos site vecters: simplified cosmid cloning. Gene 26, 137-146.

Bernard, O., Hozumi, N. and Tonegawa, S. (1978) Sequences of mouse immunoglobulin light chain genes before and after somatic changes. Cell 15, 1133-1144.

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

Boothroyd, J. C., Paynter, C. A., Coleman, S. L. and Cross G. A. M. (1982). Complete nucleotide sequence of complementary DNA coding for a variant surface glycoprotein from Trypanosoma brucei. J. Mol. Biol. 157, 547-556.

Breathenach, R., Benoist, C., O'Hare, K., Ganon, F. and Chambon, P. (1978) Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at exon-intron boundaries. Proc. Natl. Acad. Sci. U.S.A. 75, 4853-4857.

Campbell, D. G., Gagnon, J., Reid, K. B. M. and Williams, A. F. (1981). Rat brain Thy-1 glycoprotein: the amino acid sequence, disulphide bonds

and an unuaual hydrophobic region. Biochem. J. 195, 15-30.

Chirgwin, J. M. Przybyla, A. E., MacDonald, R. J. and Rutter, W. J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18, 5294-5299.

Chou, P. Y. and Fasman, G. D. (1974) Prediction of protein conformation. Biochemistry 13, 222-245.

Cohen, F. E., Novotny, J., Sternberg, J. E., Campbell, D. G. and Williams, A. F. (1981) Analysis of structural similarities between brain Thy-1 antigen and immunoglobulin domains. Biochem. J. 195, 31-40.

Cotmore, S. F., Crowhurst, S. A. and Waterfield, M. D. (1981). Purification of Thy-1-related glycoproteins from human brain and fibroblasts: comparisons between these molecules and murine glycoproteins carrying Thy-1.1 and Thy-1.2. Eur. J. Immunol. 11, 597-603.

Cordell, B., Bell, G., Tischer, E., DeNoto, F. M., Ullrich, A., Pictet, R., Rutter, W. J., and Goodman, H. M. (1979). Isolation and characterization of a cloned rat insulin gene. Cell 18, 533-543.

Dandekar, A. M. and Qasba, P. K. (1981) Rat α-lactalbumin has a 17-residue-long COOH-terminal hydrophobic extension as judged by sequence analysis of the cDNA clones. Proc. Nalt. Acad. Sci. U.S.A. 78, 4853-4857.

Dalchau, R. and Fabre, J. W. (1979) Identification and unusual tissue distribution of the canine and human homologues of Thy-1 (θ). J. Exp. Med. 149, 576-582.

Douglas, T. C. (1972) Occurrence of a theta-like antigen in rats. J. Exp. Med. 136, 1054-1062.

Hunt, S. V., Mason, D. W. and Williams, A. F. (1977) In rat bone marrow Thy-1 anigen is present on cells with membrane immunoglobulin and on precursors of peripheral B lymphocytes. Eur. J. Immunol. 7, 817-823.

Korman, A. J., Auffray, C., Schamboeck, A and Strominger, J. L. (1982) The amino acid sequence and gene organization of the heavy chain of the HLA-DR: homology to immunoglobulins. Proc. Natl. Acad. Sci. U.S.A. 79, 6013-6017.

Kuchel, P. W., Campbell, D. G., Barclay, A. N. and Williams, A. F. (1978). Molecular weights of the Thy-1 glycoproteins from rat thymus and brain in the presence and absence of Deoxycholate. Biochem. J. 169, 411-417.

Larhammar, D., Schenning, L., Gustafsson, K., Wiman, K., Claesson, L., Rask, L. and Peterson, P. A. (1982) Complete amino acid sequence of an HLA-DR antigen-like β chain as predicted from the nucleotide sequence: similarities with immunoglobuin and HLA-A, -B and -C antigens. Proc. Natl. Acad. Sci. U.S.A. 79, 3687-3691.

Ledbetter, J. A. and Herzenbrg, L. A. (1979) Xenogeneic monoclonal antibodies to lymphoid differentiation antigens. Immunol. Rev. 47, 63-90.

Lehrach, H., Diamond, D., Wozney, J. M. and Boedtker, H. (1977) RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. Biochemistry 16, 4734-4742.

Lennon V. A., Unger, M. and Dulbecco, R. (1978) Thy-1: a differentiation marker of potential mammary myoepithelial cells in vitro. Proc. Natl. Acad. Sci. U.S.A. 75, 6093-6097.

Letarte-Muirhead, M., Barclay, A. N. and Williams, A. F. (1975) Purification of the Thy-1 molecule, a major cell surface glycoprotein of rat thymocytes. Biochem. J. 151, 685-697.

Lingappa, V. R., Katz, F. N., Lodish, H. F. and Blobel, G. (1978) A signal sequence for insertion of a transmembrane glycoprotein. J. Biol. Chem. 253, 8667-8670.

Maxam, A. M. and Gilbert, W. (1977) A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. 74, 560-564.

Maniatis, T., Fritsch, E. and Sambrook, J. (1982) Molecular cloning. (Cool Spring Harbor Laboratory, New York).

Mansour, M. H. and Cooper, E. L. (1984) Purification and characterization of Rana Pipiens brain Thy-1 glycoprotein. J. Immunol. 132, 2515-2523.

McKenzie, J. L. and Fabre, J. W. (1981) Studies with a monoclonal antibody on the distribution of Thy-1 in the lymphoid and extracellular connctive tissues of the dog. Transplantation 31, 275-282.

Moriuchi, T., Chang, H-C., Denome, R. and Silver, J. (1983) Thy-1 cDNA sequence suggests a novel regulatory mechanism. Nature 301, 80-82.

Murayama, J-I., Tomita, M. and Hamada, A. (1982) Primary structure of horse erythrocyte glycophorin HA: its amino acid sequence has unique homology with those of human and porcine erythrocyte glycophorins. J. Membrane Biol. 64, 205-215.

Nevins, J. R. (1983) The pathway of eukaryotic mRNA formation. Ann. Rev. Biochem. 52, 441-466.

Owen, M., Knott, J. C. A. and Crumpton, M. J. (1980) Labeling of lymphocyte surface antigens by the lipophilic photoactivatable reagent Hexanoylidiiodo-N-(4-azido-2-nitrophenyl)Tyramine. Biochemistry 19, 3092-3099.

Reif, A. E. and Allen, J. M. (1964) The AKR thymic antigen and its distribution in leukemias and nervous tissues. J. Exp. Med. 120, 413-433.

Rostas, J. A. P., Shevenan, T. A., Sinclair C. M. and Jeffrey, P. L. (1983) The purification and charaterization of a Thy-1-like glycoprotein from chicken brain. Biochem. J. 213, 143-152.

Scheid, M., Boyse, E. A., Carswell, E. A. and Old, L. J. (1972) Serologically demonstrable alloantigens of mouse epidermal cells. J. Exp. Med. 135, 938-955.

Seki, T., Chang, H-C., Moriuchi, T., Denome, R., Proegh, R. and Silver, J. Structural organization of the Thy-1 genes. Science (submitted).

Stern, P. L. (1973) Alloantigen on mouse and rat fibroblasts. Nature 246, 76-78.

Trowbridge, I. S., Weissman, I. L. and Bevan, M. J. (1975) Mouse T-cell surface glycoprotein recognized by heterologous anti-thymocyte sera and its relationship to Thy-1 antigen. Nature 256, 652-654.

Tyler, B. M., Cowman, A. F., Gerondakis, S. D., Adams, J. M. and Bernard, O. (1982) mRNA for surface immunoglobulin α chains encodes a highly conserved transmembrane sequence and a 28-residue intracellular domain. Proc. Natl. Acad. Sci. U.S.A. 79, 2008-2012.

Williams, A. F. and Gagnon, J. (1982) Neural cell Thy-1 glycoprotein: Homology with immunoglobulin. Science 216, 696-703.

CLOSING STATEMENT

Although many of the questions concerning the structure, polymorphism, and gene organization of class II genes have been answered by recombinant DNA technology and monoclonal antibodies analysis, there are still a set of questions that need to be answered. These include the detailed tertiary structure, the mechanism of biosynthesis, the mechanism of generation of polymorphism, the molecular function and the evolutionary history of class II antigens.

It is the same situation for Thy-1. Although the detail genomic structures of mouse, rat and human Thy-1 have been studied, many controversies concerning the Thy-1 antigen remain to be resolved. Such as, why is human Thy-1 antigen not conserved by evolution; whether the Thy-1 antigen contains the hydrophobic segment on its carboxyl end while spanning on the membrane. Finally, since Thy-1 is expressed in various amounts in different tissues, we would like to know more about how its expression is regulated, and what its role is in these different tissues.