**Transcriptional Regulation of T Cell Receptor Genes**

**by a Novel CACCC Box Binding Protein**

Thesis by

Yukang Wang

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1993

(Submitted July 20, 1992)

To my beloved country

**Acknowledgements**

I first would like to thank Dr. Leroy Hood, my advisor, for his advice, financial and emotional support, and for the opportunity to carry out independent research. I also thank the other members of my thesis committee, Drs. Barbara Wold, Kai Zinn, Carl Parker and James Strauss. I am very grateful to Dr. Joan Kobori for being my mentor on molecular biology, for valueable advice on my project, and for insightful comments on all of my manuscripts. I want to thank Drs. Joan Kobori, Sally Orr, Bruce Birren, Carol Readhead and Ms. Caryn Tong for their friendship and encouragement, and for making my graduate study at Caltech an unforgettable time of my life. I am grateful to Karen Thompson, Bertha Jones, Jessie Walker, Maria DeBruyn, Elaine Gese and Chin-Sook Kim for making the lab a very pleasant place to work. I thank Wenrong Qiu, David Wang, Kangsheng Wang, K.C. Cheng, and Ebrahim Zandi for useful discussions and for solving technical problems in my experiments. I also thank Dr. Jeffrey Leiden of the University of Chicago for critical comments on my results. I am grateful to K.C. Wong Education Foundation Inc., Ltd., Hong Kong for supporting the first six months of my graduate education. Finally, I would like to thank my wife, Li Li, for her support and patience.

## Abstract

The vertebrate immune response consists of humoral and cellular immune reactions, which are mediated mainly by immunoglobulins (Ig) and Tcell receptors (TCR) respectively. The organization of Ig and TCR genes has been well established. Each of the Ig and TCR genes consists of multiple germ line gene segments that rearrange during lymphocyte development to generate diverse receptor structures expressed on mature B and T cells. The transcriptional regulation of Ig genes has been well studied. The octamer motif in the Ig gene promoter or enhancer, the E-box and the κB site have been functionally characterized. The regulation of transcription factors that bind to these sites is well understood. The transcriptional regulation of TCR genes is not as well studied as that of Ig genes. TCR-α, -β, -γ and -δ gene enhancers and a TCR α gene silencer have been reported. Some of the transcription factors that bind to these cis-elements have been cloned. A T cell-specific transcription factor, GATA-3, may play an important regulatory role on the expression of TCR genes in T cells. The promoters of TCR genes also have been investigated, however, the transcription factors that interact with them have not been characterized. The aim of this thesis was to isolate and characterize transcription factors that function in TCR gene transcription.

A cDNA clone htβ, encoding a zinc finger protein that binds to the promoter region of the human TCR gene Vβ8.1, was cloned from a human peripheral blood T cell library. The region of this protein containing four zinc fingers of the class $Cys_2$-$X_{12}$-$His_2$ may be responsible for DNA binding to the TCR Vβ8.1 promoter sequence GAAGTTGGGGGTGGTG. A putative transcriptional activation domain that is highly negatively charged has also been found in htβ. Analysis of expression of htβ mRNA reveales similar expression levels in Hela cells, Jurkat T cells, Ramos B cells and U-937 monocyte line. In addition to binding to the human TCR Vβ8.1 promoter, htβ also can bind to the mouse TCR gene α silencer. The comparison of htβ binding sites between the human TCR Vβ8.1 promoter and the mouse TCR gene α silencer reveals a core

sequence of the CACCC box. Gel-shift assay analysis of five repeats of the CACCC box with bacterially expressed htβ protein indicates that htβ can bind to the CACCC box. Gel-shift assays of the CACCC box with nuclear extracts from various cell lines reveal four common bands in T cell, B cell, monocyte and Hela cell lines, and one extra band in Hela cell extracts. CAT assay analysis indicates the CACCC box is essential for efficient transcription of the Vβ8.1 promoter. Cotransfection with a htβ expression plasmid and a reporter plasmid show that htβ can activate human TCR Vβ8.1 gene transcription. Htβ also is able to counteract the silencing effect of the TCR α silencer. Htβ may have an interaction with the cAMP response element binding protein (CREB) to negatively regulate human Vβ8.1 gene transcription in Hela cells, and that negative effect is not significant in Jurkat T cells. The CACCC box has been found in almost all Vβ8 subfamily members ( 4 of 5 Vβ8 members in human, and 2 of 3 Vβ8 members in mouse), and both TCR α and β enhancers in human and mouse. These results suggest that the CACCC box binding protein may have an important function in the immune system.

A murine zinc finger protein (M-zif) has been isolated and characterized. It has four fingers in the zinc finger domain, the putative DNA binding domain. Also, a glutamine-rich region was found, which may be involved in transcriptional activation. A previously reported cDNA molecule was shown to contain in opposite orientatioins the coding regions of both the interleukin-2 receptor α (IL-2Rα) and M-zif genes. The results presented here indicate that the cDNA is a chimeric molecule resulting from cloning artifact. The zinc finger domain of M-zif is highly homologous to that of htβ, a human T cell receptor Vβ8.1 promoter binding protein. They may have a similar DNA binding site. M-zif is not the mouse equivalent of htβ.

## Table of Contents

**Introduction**.....1

**Chapter 1** Characterization of a cDNA Encoding a Zinc Finger Protein That Binds to the Human T Cell Receptor Gene Vβ8.1 Promoter.....34

**Chapter 2** Transcriptional Regulation of T Cell Receptor Genes Mediated by a Novel CACCC Box Binding Protein Htβ.....65

**Chapter 3** Isolation of a Mouse cDNA That is Homologous to the Human T Cell Receptor Gene Vβ8.1 Promoter Binding Protein Htβ at Zinc Finger Region.....108

**Conclusion**.....135

# INTRODUCTION

## I. TRANSCRIPTION OVERVIEW

Transcription in eukaryotes is governed by three different RNA polymerases, termed RNA polymerase I, II and III. Each class of RNA polymerase has its special functions. RNA polymerase I transcribes ribosomal RNA, RNA polymerase II transcribes messenger RNA, and RNA polymerase III transcribes tRNAs and other small RNAs (1-3). The transcription of an eukaryotic gene by RNA polymerase is governed by cis-acting elements: a promoter, an enhancer and sometimes a silencer (4-5). A promoter is a region of DNA sequence that is required for initiation of transcription. It includes the start site of transcription, often has a TATA box and a set of upstream elements. An enhancer is a DNA element located in or around a gene that increases the level of transcription, independent of its position or orientation. A silencer has the basic features of an enhancer, but decreases the level of transcription.

Transcription factors bind and recognize specific sequences within cis-acting elements to regulate gene expression. These proteins have at least two basic domains, one that recognizes a specific DNA sequence motif in a given promoter or enhancer and a second domain that interacts with other transcription factors or RNA polymerase to regulate the level of transcription (6-7). On the basis of protein structural motifs at the DNA binding domains, transcription factors can be divided into four groups: helix-turn-helix, zinc finger, leucine zipper and helix-loop-helix (fig. 1) (6-7).

### A. DNA binding domains

#### 1). Helix-turn-helix

The first characterized structural motif of a DNA-binding domain is the helix-turn-helix found in prokaryotic activator and repressor proteins (8). The basic structure

of this class of proteins is two α-helices that are separated by a β-turn. In eukaryotic organisms the helix-turn-helix motif first was found in several regulatory proteins encoded by homeotic genes of *Drosophila* embryogenesis (9-10). A conservative DNA segment of approximately 180 bp in length is characteristic for these homeotic genes, termed homeobox. Sequences homologous to the homeobox have been isolated from higher organisms, such as mammals (11). The homeobox proteins are characterized by a helix-turn-helix motif in which one of the helices called the "recognition helix" directly binds to specific DNA sequences. The N-terminal helix is 8 amino acids long, the turn is 3 amino acids long, and the C-terminal helix is 9 amino acids long. The homeobox proteins are capable of sequence-specific DNA binding in vitro (12). In cultured *Drosophila* cells and in a transient expression system, the homeobox proteins have been shown to either stimulate or repress transcription of certain genes (13-14). They are transcriptional regulators active during development. Also there is a subclass of transcription factors within the homeobox protein family that has a common bipartite domain, termed POU domain, which was first described in the mammalian PIT-1, OCT-1, OCT-2, and *C. elegans unc-86* gene products (15). The POU domain contains a homeodomain and a second region, the POU box. The POU box is unique to this subclass.

**2). Zinc finger**

Zinc finger motifs were defined first as a DNA-binding structure in the RNA polymerase III transcription factor TFIIIA that is involved in the expression of the *Xenopus laevis* 5S ribosomal RNA gene (16). Several zinc fingers containing proteins have been described that function in transcription mediated by RNA polymerase II (17). These zinc fingers containing transcription factors can be classified as two types. The first type is TFIIIA-like zinc fingers that consist of two cysteine and two histidine

residues that form a domain by coordinating a $Zn^{2+}$ ion; there is a region of approximately 12 amino acids between zinc fingers. The sequence in this region is quite conservative among species. This type of zinc finger motif has been found in a variety of eukaryotes, notably the mammalian transcription factor Sp1 (18-19) and *Drosophila Kruppel* proteins (20). The zinc finger region and zinc are necessary for the DNA binding function of these transcription factors. The second class of zinc finger motif has two pairs of cysteines. Mutations at the key cysteine residues result in elimination of sequence-specific DNA-binding function (21). This class of zinc finger motif has been observed in a variety of DNA-binding proteins, from mammalian hormone receptors (22-23) to yeast regulatory proteins (24). The zinc finger containing proteins exhibit many DNA sequence specificities. The highly conserved amino acids in the finger region are not responsible for the DNA binding specificities, but may provide a common structure for the binding domain. The crystal structure of a zinc finger protein, mouse immediate early protein Zif268, suggests that some of the nonconserved amino acid residues determine the specificity (25).

**3). Leucine zipper**

The third class of DNA binding protein is called leucine zipper, which was first described for the rat liver enhancer binding protein c/EBP (26). The sequence similarity among c/EBP and several oncogene products (Fos, Myc, and Jun), and also a yeast regulatory protein (GCN4) was noted. The common structure of these transcription factors is a region of highly conserved sequence with positively charged amino acids immediately followed by a region of a strict heptad repeat of leucine residues, named "leucine zipper" (26). The leucine zipper is required for dimerization of proteins in this group. The basic region is necessary for DNA binding, and the leucine zipper region is believed to assist this basic region to form a DNA binding structure (26). Other

proteins, for example, OCT-2 contains a leucine zipper without the conserved basic region, and is believed to use its homeodomain for specific DNA binding (27).

#### 4). Helix-loop-helix

The fourth class of DNA binding motif termed helix-loop-helix was identified in several regulatory proteins, the *daughterless* and *achete scute* gene products in *Drosophila* (28-29), the Myc, MyoD, and two immunoglobulin enhancer binding proteins in mammals (30-31). They all contain a conserved basic region followed by a motif that consists of two amphipathic helices separated by a loop of variable length. The basic and helix-loop-helix (HLH) regions allow these proteins to dimerize with themselves and/or with other members of the family, and bind to DNA (32-34). The HLH motif is mainly for dimerization, while the basic region is responsible for DNA binding and functional specificity (34). The HLH proteins lacking a basic region may regulate negatively other HLH proteins by forming a nonfunctional heterodimer (35). Besides these four classes of structural motifs (helix-turn-helix, zinc finger, leucine zipper and HLH), there are other transcription factors, for example heat-shock transcription factors (36-37), CTF/NF1 (38), HAP2 and HAP3 (39-40), and TFIID (41-42), which fall outside the four classes. It is anticipitated that more new DNA binding domain motifs will be identified.

### B. Transcriptional activation domains

The transcriptional regulation function of DNA binding proteins is dependent upon a region of sequence that is separated from the DNA binding domain. One transcription factor can have more than one transcriptional activation domain and can be interchangeable. When the transcriptional activation domain from one factor is linked to

the DNA binding domain of another factor, the chimeric transcription factor retains the DNA binding specificity and transcriptional activation function of the original transcription factor. Transcriptional activation domains can be categorized into three classes: acidic domains, glutamine-rich domains and proline-rich domains (fig. 2) (43).

**1). Acidic domain**

The transcription activation domains were first identified in the yeast regulator proteins GAL4 and GCN4 (44-46). The activation domains of these two factors consist of a region of amino acids with two common features. They are highly negatively charged and can form an amphipathic α-helix, although they do not have obvious sequence similarity. When the acidic domain is linked to a different DNA binding domain, it can activate the transcription of a reporter gene according to the DNA binding domain specificity. This has been shown in yeast and in cells of many kinds of higher organisms (44). This acidic domain has been found in mammalian transcription factors AP-1/Jun and TEF-3 and correlates with their transcriptional activation function (47-48). The acidic domain may interact with transcription initiation components, very likely the TATA box binding protein TFIID and works universally on gene transcription (42, 49). The evidence is that the herpes simplex virion protein VP16 that has a strong acidic domain binds to both human and yeast TFIID selectively (50).

**2). Glutamine-rich domain**

The second class of transcriptional activation domains was found in the human Sp1 factor (51). Sp1 has two strong activators containing approximately 25% glutamine and few charged amino acid residues. This glutamine-rich motif also was found in *Drosophila antennapedia ultrabithorax* (51), in yeast HAP1 and HAP2 (52-53), and in

mammalian OCT-1, OCT-2, TFIID and Jun (54-56, 47,42). There are no obvious sequence homologies among these glutamine-rich regions. The glutamine-rich region from one transcription factor can be linked to the DNA binding domain from another transcription factor and still have activation function (51). The glutamine-rich activator may work only in certain cell types, for example, it has not been shown that the glutamine-rich domains from Sp1 can activate transcription in yeast. As proposed, the glutamine-rich domain may either bind to an acidic activator that is attached to TFIID, or contact a part of TFIID different from that recognized by an acidic activator, or interact with another component of the transcriptional initiation complex other than TFIID (49). By using one of these three mechanisms, glutamine-rich domains can activate gene expression.

**3). Proline-rich domain**

The third class of activation domain was defined in CTF/NF-I (43). The proline-rich domain of CTF can activate transcription when it forms a chimeric transcription factor with other DNA binding domains. Besides CTF, there are many mammalian transcription factors containing a proline-rich domain, for example Jun and OCT-2 (48,55-56).

All three classes of transcriptional activation domains are likely to carry out their functions by interacting with other proteins. Although acidic domains interact with the TATA box binding protein TFIID (50), whether the glutamine-rich and proline-rich domains contact different parts of TFIID is unknown. It is possible that both glutamine-rich and proline-rich domains interact with other components in the transcription initiation complex and activate transcription.

## II. OVERVIEW OF IMMUNOGLOBULIN AND T CELL RECEPTOR GENE STRUCTURE

The vertebrate immune response consists of humoral and cellular immune reaction against foreign invaders and abnormal self-proteins. The humoral immune response is mainly mediated by B lymphocytes and their secreted molecule, an immunoglobulin (Ig) that recognizes antigen alone. The cellular immune response largely depends on T lymphocytes that bear T cell receptors (TCR) on their surface that recognize antigen peptide in association with class I and class II major histocompatibility complex (MHC) proteins, termed MHC restriction (57). T cells acquire MHC restriction during their maturation in the thymus (58). Mature T cells express either $\alpha\beta$ or $\gamma\delta$ TCR molecules on their surface. T cells have at least two subclasses, helper T cells that recognize foreign antigen in association with class II MHC molecules and cytotoxic T cells that recognize foreign antigen and the class I MHC molecule. Besides TCR, there are a number of accessory molecules playing important roles in T cell recognition and activation, notably, CD3, CD4 and CD8. Usually helper T cells express CD3 and CD4, and cytotoxic T cells express CD3 and CD8. The CD3 molecule consists of $\gamma$, $\delta$, $\varepsilon$, $\zeta$ and $\eta$ chains (59).

The structure of Ig and TCR genes is known (fig. 3). The Ig genes have three families, the heavy chain family and the $\kappa$ and $\lambda$ light chain gene families (60). Every variable gene is composed of two parts, a variable gene segment (V) and a constant gene segment (C) joined through a DNA rearrangement process. A $V_L$ gene is produced by combining a V gene segment and a joining ($J_L$) gene segment, while a $V_H$ gene is constructed from three gene segments, a V, a diversity (D) and a J gene segment. Then the $V_LJ_L$ or $V_HD_HJ_H$ joins the $C_\kappa$, $C_\lambda$ or $C_H$ gene segment via RNA splicing. A similar rearrangement process occurs with TCR genes. V$\alpha$ and V$\gamma$ genes are generated by V and J gene segment joining; V$\beta$ and V$\delta$ genes are formed by combining V, D, and

J gene segment (61-62). The TCR δ genes are located within the TCR α locus (63). The $C_\delta$ gene is located 5' of the Jα gene segments, and both Vα and Vδ may rearrange to it. The variable gene segment rearrangement is mediated by rearrangement signals located 3' to the V gene segments, 5' to the J gene segments, and on both sides of the D gene segments (62). The rearrangement signal consists of a conserved heptamer, 5' CACAGTG 3' and an A/T rich nanomer, separated by a spacer sequence of either 12 or 23 base pairs (64-65). Usually the gene segments with different spacer length can join each other (63-64).

Transfected TCR genes can rearrange in B cell lines (66). Hence, the VDJ joining of TCR and Ig genes may use a common recombination machinery. Both RAG-1 and RAG-2 gene products are capable of providing fibroblasts with the competence to carry out VDJ joining. These two genes are expressed in immature B and T cells, but not in mature cells of either kind (67-68). Both RAG-1 and RAG-2 deficient mice have no mature B and T lymphocytes. The arrest of B and T cell differentiation occurs at an early stage and results from the inability of initiating V(D)J rearrangement (69-70). The exact functions carried out by RAG-1 and RAG-2 in gene rearrangement are not clear.

Generating diversity in the antigen receptor is necessary to accommodate the large number of antigens in the environment. Both the Ig and TCR genes use similar mechanisms for diversification during and after gene rearrangement. The first mechanism is germline diversity. Both Ig and TCR genes have a large number of different gene segments for forming their variable region genes. The combinatorial joining of these gene segments is the second mechanism, and during the joining process, nucleotides can be added or deleted to the joining ends of gene segments. In addition, there are multiple translational reading frames in the D gene segments (62). After rearrangement, Ig genes employ the third mechanism, termed somatic hypermutation, for generating diversity. TCR genes rarely or do not use this mechanism. Somatic

hypermutation,the acquisition of specific single nucleotide change, can happen throughout the V gene and its flanking sequences to generate higher affinity immunoglobulin molecules, which is the mechanism of affinity maturation in a secondary immune response (62). By using these mechanisms the immune system can produce enough antigen receptors to recognize all kinds of antigens in the environment (62, 71).

Ig gene rearrangement and expression have four separate stages (72). The first stage is $D_H$-$J_H$ rearrangement, followed by the $V_H$ gene segment joining to the $D_H$-$J_H$ gene segment. The third step is Vκ and Jκ joining that is triggered by heavy chain rearrangement. Finally if the κ gene has rearranged nonproductively, the λ light chain genes rearrange. TCR gene rearrangement and expression can be divided into four steps. In the fetal thymus, TCR δ and γ gene segments rearrange first, followed by Dβ-J β joining. Vβ gene segments then rearrange to Dβ-Jβ to form Vβ-Dβ-Jβ. Finally, the Vα and Jα gene segments rearrange to generate Vα-Jα (73). There are two different lineages of T cell development, the TCR αβ lineage and TCR γδ lineage (74). The TCR γδ heterodimers appear before αβ in fetal thymocytes. T cell precursors first rearrange γ and δ genes, and those that succeed become γδ T cells. Those that failed to produce γ and δ chains may then rearrange TCR β and α loci, and differentiate into αβ T cells (75).

## III. TRANSCRIPTIONAL REGULATION OF IMMUNOGLOBULIN AND T CELL RECEPTOR GENES

The mammalian bone marrow is one of the best studied model systems of cell differentiation. A number of cell lineages are generated from bone marrow pluripotent stem cells: the megakaryocyte-platelet lineage, the erythroid lineage, the granuolocyte-

monocyte lineage and the lymphoid lineage. Both T and B cells are differentiated from precursor cells of lympholineage. T and B cell development provide an excellent system to study tissue-specific gene expression, and to identify tissue-specific transcripton factors.

**A. Transcriptional regulation of immunoglobulin gene expression**

Ig gene transcription has been used as a model system to understand B cell-specific gene expression for many years (Table 1). An octamer motif 5' ATTTGCAT 3' or its inversion has been found in all Ig promoters (76). Synthetic promoters with the octamer and a TATA box are active only in lymphoid cells, suggesting that the octamer motif is lymphoid-restricted (77). However, in some cases, the octamer activates transcription in non-B cell types (78-79). The phenomenon was not understood until the cloning of OCT-1 and OCT-2 transcription factors (54-56). Both OCT-1 and OCT-2 can bind to the octamer, and are members of the POU-domain family (15). OCT-1 is expressed in a variety of cell types, while OCT-2 is limited to B cells (54-56). This suggests that OCT-1 works in a general way and OCT-2 mediates lymphocyte-specific activity. Additional transcription factors, OCT-3 and PIN-1, have been identified that also interact with the octamer (80-82).

The Ig heavy chain (IgH) gene enhancer was the first cellular enhancer to be identified. It is located in the intron between the J and C gene segments. This enhancer is B cell-specific (83). B cell-specific binding activities have been observed to the octamer motif. Both OCT-1 and OCT-2 can bind to the octamer motif in the enhancer (54-56). The second element that is active only in B cells is the μB motif (84-86). A set of homologous motifs in the enhancer is termed the E motifs (87). Molecular cloning of cDNAs for the E motif binding proteins has identified a new family of transcription factors, the helix-loop-helix proteins (32). The HLH proteins that bind to

the E motifs can be divided into two groups, one group specific for μE5 of the IgH enhancer and κE2 of the Igκ enhancer and another group specific for the E3 motif in both the IgH and Igκ enhancers. The μE5/κE2 binding proteins include E12, E47, ITF-1 and ITF-2. E12, E47 and ITF-1 are derived by alternative splicing from the same gene located on human chromosome 19, and ITF-2 is encoded by a different gene (32, 88-90). This group of proteins includes the basic motif and HLH structure (bHLH). The μE3/κE3 binding proteins have TEF-3, TFEB and USF (or NF-μE3) and share a dimerization domain leucine zipper structure besides the basic motif and HLH structure (bHLH-zip) (91-92). The 60 base pair region of the IgH enhancer that binds bHLH and bHLH-zip proteins is sufficient for B cell-specific transcription (93). ITF-1 and TEF-3 can directly activate transcription when they bind to their cognate binding sites within the enhancer region of a gene (88,91). TEF-3 has an α-helical and negatively charged region that is the transcription activation domain (91). In non-B cells, there may be negative regulation restricting HLH protein-mediated activation. It has been proposed that a repressor in non-B cells binds the μE5 site, and causes inactivation of the μE3 site, since overexpression of ITF-1 in fibroblasts can activate transcription through the μ E5 site and relieves repression. In B cells, the μE5-binding repressor may be absent or ITF-1 has a much higher binding affinity in B cells than in fibroblasts because of different post-translational modification of ITF-1 in different cell types (91-94). The Id protein that has the HLH structure but does not contain the basic amino acid region can bind to bHLH protein to form a heterodimeric complex, leading to inactivation of bHLH protein (35). It has been shown that Id inhibits the activity of both IgH and Igκ enhancers in B cells (95).

The Igκ enhancer also is located in the intron between the J and C gene segments. It has at least two important motifs, κB and E. The transcription factor NF-κ B recognizes the 10-bp κB motif that is crucial for enhancer function (83). NF-κB binding activity is only detected in mature B cells, but can be induced in pre-B cells

(96). NF-κB contains two subunits of about 50 and 65 kd (97). The smaller subunit contains the DNA binding function (98-99), while the bigger subunit does not bind DNA but is essential for inhibition of NF-κB by I-κB that is a 35 kd inhibitor protein (97, 100-101). Usually cytoplasmic NF-κB is associated with I-κB in a latent form. Upon induction, protein kinases activate NF-κB by phosphorylation and thereby dissociate of I-κB (101). NF-κB translates into the nucleus, binding to its cognate site and activating transcription. Many important genes in immune function can be induced by NF-κB, including IL-2, IL-2 receptor α chain, IL-6, TNF-α and β, MHC class I and II genes, and β-interferon (102). Since NF-κB can be induced in a variety of cells, how mature B cells keep NF-κB in an active form constitutively is a very interesting question. Another motif in the Igκ enhancer is the E box. E protein heterodimers that bind to μE5 and μE3 in the IgH enhancer interact with κE2 and κE3 sites respectively. Functional cooperation between E motifs seems very important for enhancer function since mutations at a κE site affect enhancer activity dramatically (103). In addition, two B cell-specific enhancers that are located downstream of the $C_H$ gene and $C_\kappa$ gene have been found (104-105). The IgH enhancer has an octamer motif and a μE5 motif, and the Igκ enhancer has two potential NF-κB sites and a μB motif. Interestingly, the Igκ enhancer is very strong even in the absence of NF-κB. The detailed regulatory functions of these enhancers are not quite clear.

## B. Transcriptional regulation of T cell receptor genes

### 1). Transcription in the TCR α gene

The promoter of a rearranged TCR α gene is T cell-specific when tested in a CAT assay in which a TCR α promoter and CAT gene containing plasmid was transfected into various cell lines (106). A T cell-specific enhancer was found several

kilobases downstream from Cα in both human (107) and mouse (108). These two enhancers have 120 base pairs of homology. At least four proteins interact with the mouse enhancer. Two of these proteins are not tissue-specific, and the other two proteins are T cell-specific (108). In the human TCR α enhancer, four functional nuclear protein binding sites, Tα1-Tα4 have been identified. The minimal TCR α enhancer is contained within the Tα1 and Tα2 sites (109). The spacing between Tα1 and Tα2 is critical for enhancer function; 20-25 bp spacing is adequate for enhancer activity (110). Tα3 and Tα4 may represent redundant elements, because they compensate for mutations in Tα1 or Tα2 that abolish enhancer activity (109).

Several transcription factors regulate the activity of the TCR α enhancer (Table 2). Three cAMP response element binding proteins (CREB) bind to Tα1 that contains a cAMP response element (TGACGTCA) (107, 110-111). Tα2 has a pyrimidine-rich sequence (CCCTTTGAAG) at its 5' end and another pyrimidine-rich sequence (CATCCTCT) at its 3' end. Two transcription factors TCF-1 (112) and TCF-1α (113) in the human bind to the 5' end pyrimidine-rich sequence (Table 2). LEF-1 seems to be the mouse homolog of the human TCF-1α (114). These proteins contain a DNA-binding domain that has high homology to the DNA binding domain of proteins in the high mobility group (HMG) family. LEF-1 is expressed in T cells of all stages and mouse B cells. TCF is expressed only in T cells. The binding protein for the 3' end pyrimidine-rich region of Tα2 is the Ets proto-oncogene product (115), which is expressed in T and B cells (116). Mutation in the core Ets-1 binding site in Tα2 abolishs the minimal α TCR enhancer (115). Just like other Ets family members, Ets-1 contains a conserved basic domain and an adjacent α-helical region (117). Both regions are necessary for DNA binding (110). Tα3 has the consensus binding site (AGATAG) for GATA zinc finger proteins (109). The GATA-3 gene encoding the GATA protein has been cloned from human (118), mouse (119), and chicken (120). Human GATA-3 is expressed in T

cells, kidney and brain. Mouse GATA-3 is expressed in early thymocytes (110), suggesting GATA-3 is a good candidate for a T cell-determining gene.

In mouse, a negative regulatory element has been identified near the TCR α enhancer, which is called the TCR α silencer (121). The TCR α silencer can down regulate transcription from promoters in γδ but not in αβ T cells by silencing the α enhancer activity. The lineage specificity of the α enhancer is achieved through the silencing function.

**2). Transcription in the TCR β gene**

Tissue specificity of the Vβ promoter has been addressed by several studies. The Vβ promoter only functions in T cells, not in other cells (122). Sequence analysis of 14 murine Vβ promoters show a conserved decamer motif (AGTGAT/CG/ATCA) 10-40 bp upstream of the TATA box. This decamer has the cyclic AMP response element (CRE) (123). There is a nuclear protein that can specifically recognize the Vβ promoter region (124). The decamer sequence was present also in a human Vβ promoter. Nuclear extracts from various human cell lines can bind to this Vβ decamer (125). It is unlikely that the decamer binding protein is solely responsible for tissue specificity of Vβ promoters. In another study, within 175 bp upstream of the transcription initiation site of the human Vβ8.1 gene, there are four protected regions, at nucleotide -46 to -68 (I), -72 to -92 (II), -113 to-134 (III), and -136 to -175 (IV). Nuclear proteins of a variety of cell types produced footprints I, III, and IV. Footprint II was produced only by T cell extracts, which may be the binding site of T cell-specific factors (125). In agreement with the above result, the Vβ promoter was only functional in T cells (122).

The murine and human T cell receptor gene β enhancer is located 3' of Cβ2 (126-128). Within this enhancer region, a number of enhancer motifs have been noted. Some of the transcription factors interacting with those motifs have been characterized

(Table 2). Two DNase I hypersensitive sites have been described downstream of Cβ2 (129), one site overlapping with the TCR β enhancer and the other was expressed only in TCR β-expressing T cells, and is suggested to have a stage-specific regulatory function in T cell development.

### 3). T cell-specific transcription in the γ and δ TCR genes

Experiments with transgenic mice have shown that there is a silencer element in sequence flanking the TCR Cγ locus. Deletion of this element results in expression of the TCR γ gene in cells that otherwise express the TCR αβ heterodimer (130). A 1.8 kb fragment located 3 kb 3' to the Cγ exon III was found to display enhancing activity in T cells, but it is inactive in non-T cells (131). A T cell lineage-specific enhancer in the Jδ 3-Cδ intron also has been described. Enhancer activity has been localized to a 250 bp region that contains multiple binding sites for nuclear proteins (Table 2) (132-133). Since the TCR δ gene is located within the TCR α locus, and the TCR α gene has its own enhancer, it seems that transcriptional control of the TCR δ and α gene is mediated by distinct enhancers.

## IV. THESIS

Although the transcriptional regulation of immunoglobulin genes is well-studied, little is known about T cell receptor gene transcription. The octamer motifs are present in both the immunoglobulin gene promoter and enhancer, which are responsible for the B cell-specific expression of Ig genes. The conservative sequences specifically required for TCR gene transcription have not been well-characterized in the TCR gene promoters and enhancers. T cell-specific TCR gene expression is interesting since there are many

TCR variable genes and there may be a T cell-specific transcriptional regulation that is common to all of these variable genes. Also, T cells play key roles in the immune system. T cells recognize foreign antigen through their surface TCR gene products. Understanding T cell receptor gene expression will be helpful to understand T cell development, tolerance, autoimmunity and other major problems in the immune system. Previous studies have shown that the human TCR Vβ gene promoter is T cell specific, and that Vβ gene expression may be regulated by T cell-specific transcription factors. In my work, I use the well-characterized TCR Vβ8.1 gene promoter as a model to understand TCR gene transcriptional regulation by DNA binding proteins. My goal was to clone T-cell specific transcription factors and study their functions on TCR gene expression.

Chapter 1 of this thesis describes the isolation and characterization of a cDNA clone encoding a DNA binding protein, htβ, which recognizes a GT rich motif in the promoter region of the human TCR Vβ8.1 gene. Htβ contains a $Cys_2$-$X_{12}$-$His_2$ zinc finger domain and an acidic region that is the putative transcriptional activation domain. Htβ expression is not T cell specific.

Chapter 2 further characterizes htβ as a CACCC box binding factor. The CACCC box is essential for efficient human Vβ8.1 gene expression. Htβ activates human Vβ8.1 gene transcription and antagonizes the effect of the mouse TCR α silencer. The CACCC motif is observed in many TCR regulatory sequences and other gene regulatory regions, suggesting htβ may have a more general function on gene expression. Htβ may have an interaction with the cAMP response element binding protein (CREB).

Chapter 3 describes the cloning of a mouse zinc finger containing protein (M-zif). The zinc finger regions of the M-zif and the htβ genes are highly similar, suggesting they may have the same DNA binding site. Clearly M-zif is not the mouse counterpart of htβ. The relationship between the M-zif gene and the interleukin-2 receptor α gene is explored and discussed.

In conclusion, a gene for a CACCC box binding protein has been cloned. This CACCC box binding protein activates the human TCR Vβ8.1 gene and antagonizes the effect of the mouse TCR gene α silencer. It possibly binds to other TCR regulatory sequences, and regulates transcription of many genes other than TCR genes. This work provides the basis for understanding the regulating function on gene expression of a CACCC box binding protein. A gene encoding a mouse zinc finger protein also has been cloned, detailed analysis of its function awaits further experimentation.

REFERENCES

1. Sollner-Webb, B. and Tower, J. (1986) Annu. Rev. Biochem. 55:801-830.

2. Matsui, T., Segall, J., Weil, P.A. and Roeder, R.G. (1980) J. Biol. Chem. 255:11992-11996.

3. Geiduschek, E.P. and Toccihini-Valentini, G.P. (1988) Annu. Rev. Biochem. 57:873-914.

4. Sefling, E., Jasin ,M. and Schaffner, W. (1985) Trends Genet. 1:224-230.

5. Voss, S.D., Schlokat, U. and Gruss, P. (1986) Trends Biochem. Sci. 11:287-289.

6. Struhl, K. (1989) Trends Biochem. Sci. 14:137-140.

7. Johnson, P.F. and McKnight, S.L. (1989) Annu. Rev. Biochem. 58:799-839.

8. Pabo, C.O. and Sauer, R.T. (1984) Annu. Rev. Biochem. 52:293-321.

9. Wharton, R.P. and Ptashne, M. (1986) Trends Biochem. Sci. 11:71-73.

10. Levine, M. and Hoey, T. (1988) Cell 55:537-540.

11. Wright, C.V.V.E., Cho, K.W.Y., Oliver, G. and DeRobertis, E.M. (1989) Trends Genet. 14:52-56.

12. Desplan, C., Theis, J. and O'Farrel, P.H. (1985) Nature 318:630-635.

13. Krasnow, M.A., Saffman, E.E., Kornfeld, K. and Hogness, D.A. (1989) Cell 57:1031-1043.

14. Biggin, M.D. and Tjian, R. (1989) Cell 58:433-440.

15. Herr, W., Sturn, R.A., Clerc, R.G., Corcoran, L.M., Baltimore, D., Sharp, P.A., Ingraham, H.A., Rosenfeld, M.A., Finney, M., Ruvkun, G., and Horvitz, H.R. (1988) Genes Dev. 2:1513-1516.

16. Miller, J., MaLachlan, A.D. and Klug, A. (1985) EMBO J. 4:1609-1614.

17. Evans, R.M. and Hollenberg, S.M. (1988) Cell 52:1-3.

18. Kadonaga, J.T., Courey, A.J., Ladika, J. and Tjian, R. (1988) Science 242:1566-1570.

19. Kadonaga, J.T., Carner, K.R., Masiarz, F.R. and Tjian, R. (1987) Cell 51:1079-1090.

20. Tautz, D., Lehmann, R., Schnurch, H., Schuh, R., Seifert, E., et al. (1987) Nature 27:383-389.

21. Berg, J.M. (1985) Science 232:485-487.

22. Weinberger, C., Hollenberg, S.M., Rosenfeld, M.G. and Evans, R.M. (1985) Nature 318:670-672.

23. Miesfeld, R., Rusconi, S., Godowski, P.J., Maler, B.A. Okret, S., et al. (1986) Cell 46:389-399.

24. Laughon, A. and Gesteland, R.F. (1984) Mol. Cell Biol. 4:260-267.

25. Pavletich, N.P. and Pabo, C.O. (1991) Science 252:809-817.

26. Landschlz, W.H., Johnson, P.F. and McKnight, S.L. (1988) Science 240:1759-1764.

27. Kouzarides, T. and Ziff, E. (1988) Nature 336:646-651.

28. Dang, C.V., McGuire, M., Buckmire, M. and Lee, W.M.F. (1989) Nature 337:664-666.

29. Villares, R. and Cabrera, C.V. (1987) Cell 50:415-424.

30. Candy, M., Vassin, H., Brand, M., Tuna, R., Jan, L., and Jan, Y.N. (1988) Cell 55:1061-1067.

31. Davis, R.L., Weintraub, H. and Lassur, A.B. (1987) Cell 51:987-1000.

32. Marre, C., McCaw, P.S. and Baltimore, D. (1989) Cell 56:777-783.

33. Lassar, A.B., Buskin, J.N., Lockshon, D., Davis, R.L., Apone, S., Hauschka, S.D. and Weintraub, H. (1989) Cell 58:823-831.

34. Davis, R.L., Cheng, P.F., Lassur, A.B. and Weintraub, H. (1990) Cell 60:733-746.

35. Benezra, R., Davis, R.L., Lockshon, D., Turner, D.L. and Weintraub, H. (1990) Cell 61:49-59.

36. Wiederrcht, G., Seto, D. and Parker, C.S. (1988) Cell 54:841-853.

37. Sorger, P.K. and Pelum, H.R.B. (1988) Cell 54:855-864.

38. Santoro, C., Mermod, N., Ardrews, P.C. and Tjian, R. (1988) Nature 334:218-224.

39. Pinkham, J.L., Olesen, J.T. and Guarente, L. (1987) Mol. Cell Biol. 7:587-588.

40. Hahn, S., Pinkham, J., Wei, R., Miller, R. and Guarente, L. (1988) Mol. Cell Biol. 8:655-663.

41. Horikoshi, M., Wang, C.K., Fuji, H., Cromlish,. J.A., Weil, P.A. and Roeder, R.G. (1989) Nature 341:299-303.

42. Kao, C.C., Lieberman, P.M., Schmidt, M.C., Zhou, Q., Pei, R. and Berk, A.J. (1990) Science 248:1646-1650.

43. Mitchell, P. and Tjian, R. (1989) Science 245:371-378.

44. Ptashne, M. (1988) Nature 335:683-689.

45. Hope, I.A. and Struhl, K. (1986) Cell 46:885-894.

46. Hope, I.A., Mahadevan, S. and Struhl, K. (1988) Nature 333:635-640.

47. Bohmann, D., Bos, T.J., Admon, A., Nishimura, T., Vogt, P.K. and Tjian, R (1987) Science 238:1386-1392.

48. Struhl, K. (1988) Nature 332;649-650.

49. Ptashne, M. and Gann, A.A.F. (1990) Nature 346:329-331.

50. Stringer, K.F., Ingles,C.J. and Greenblatt, J. (1990) Nature 345:783-786.

51. Courey, A. and Tjian, R. (1988) Cell 55:887-898.

52. Pfeifer, K., Kim, K.S., Kogan, S. and Guarente, L. (1989) Cell 56:291-301.

53. Pinkham, J., Olesen, J. and Guarente, L. (1987) Mol. Cell Biol. 7:578-585.

54. Sturm, R.A., Das, G. and Herr, W. (1988) Genes Dev. 2:1582-1599.

55. Muller, M.M., Ruppert, S., Schaffner, W. and Mutthias, P. (1988) Nature 336:544-551.

56. Scheidereit, C.,Cromlish, J.A., Gerster, T., Kawakami, K., Bulmuceda, C.G., Currie, R.A. and Roeder, R.G. (1988) Nature 336:551-557.

57. Paul, W.E. (1989) Fundamental Immunology, 2nd ed. Raven, New York.

58. Bevan, M.J. (1981) Immunol. Today Nov. 216-219.

59. Ashwell, J.D. and Klausner, R.D. (1990) Annu. Rev. Immunol. 8:139-167.

60. Honjo, T. (1983) Annu. Rev. Immunol. 1:499-529.

61. Davis, M.M. (1985) Annu. Rev. Immunol. 3:537-560.

62. Kronenberg, M., Siu, G., Hood, L. and Shastri, N. (1986) Annu. Rev. Immunol. 4:529-591.

63. Chien, Y-H., Iwashima, M., Kaplan, K.B., Elliot, J.P. and Davis, M.M. (1987) Nature 327:677-682.

64. Early, P., Huang, H., Davis, M., Calame, K. and Hood, L. (1980) Cell 19:981-992.

65. Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. and Tonegawa, S. (1980) Nature 286:676-683.

66. Yancopoulus, G.D., Blackwell, T.K., Suh, H., Hood, L. and Alt, F.W. (1986) Cell 44,251-259.

67. Schatz, D.G., Oettinger, M.A. and Baltimore, D. (1989) Cell 59:1035-1048.

68. Oettinger, M.A., Schatz, D.G., Gorka, C. and Baltimore, D. (1990) Science 248:1517-1523.

69. Mombaerts, P., Iacomini, J., Johnson, R.S., Herrup, K., Tonegawa, S. and Papaioannou, V.E. (1992) Cell 68:869-877.

70. Shinkai, Y., Rathbun, G., Lam, K-P., Oltz, E.M., Stewart, V., Mendelsohn, M., Charron, J.,Datta, M., Yong, F., Stall, A.M. and Alt, F.W. (1992) Cell 68:855-867.

71. Tonegawa, S. (1983) Nature 302:575-581.

72. Alt, F.W. (1984) Nature 312:502-503.

73. Rothenberg, E.V. (1992) Adv. Immunol. In press.

74. Winoto, A. and Baltimore, D. (1989) Nature 338:430-432.

75. Davis, M.M. (1990) Annu. Rev. Biochem. 59:475-496.

76. Parslow, T.G., Bluir, D.L., Murphy, W.J. and Granner, D.K. (1984) Proc. Natl. Acad. Sci. USA 81:2650-2654.

77. Wirth, T., Staudt, L. and Baltimore, D. (1987) Nature 329:174-178.

78. Davidson, I., Promental, P., Augerenn, A., Wilderman, M., Zenke, P. and Chambon, P. (1986) Nature 323:544-548.

79. O'Hare, P. and Goding, C.R. (1988) Cell 52:435-445.

80. Lenardo, M.J., Staudt, L.,Robbin, P.,Kuang, A., Mulligan, R.C. and Baltimore, D. (1989) Science 243:544-546.

81. Bodner, M., Castrillo, J.L., Theil, L.E., Deerinck, T., Ellisman, M. and Karin, M. (1988) Cell 55:505-518.

82. Ingraham, H.A., Chen, R., Mangalam, H.J., Elsholtz, H.P., Flynn, S.E., Lin, C.R., Simmons, D.M., Swanson, L. and Rosenfeld, M.G. (1988) Cell 55:519-529.

83. Staudt, L.M. and Lenardo, M.J. (1991) Annu. Rev. Immumol. 9:373-378.

84. Araki, K., Maedu, H., Wang, J., Kitamura, D. and Watanabe, T. (1988) Cell 53:723-730.

85. Libermann, T.A., Lenardo, M. and Baltimore, D. (1990) Mol. Cell Biol. 10:3155-3162.

86. Nelsen, B., Kadesch, T. and Sen, R. (1990) Mol. Cell Biol. 10:3145-3154.

87. Ephrussi, A., Church, G.M., Tonegawa, S. and Gilbert, W. (1985) Science 227:134-146.

88. Henthorn, P., Kiledjian, M. and Kadesch, T. (1990) Science 247:467-470.

89. Marre, C., McCaw, P.S., Vaessin, H., Candy, M., Jan, L.Y., Jan, Y.N., Cabrera, C.V., Buskin, J.N., Huushia, S.D., Lassar, A.B., Weintraub, H. and Baltimore, D. (1988) Cell 58:537-544.

90. Sun, X-H. and Baltimore, D. (1991) Cell 64:459-470.

91. Beckmann, H., Su, L-K. and Kadesch, T. (1990) Genes Dev. 4:167-179.

92. Carr, C.S. and Sharp, P.A. (1990) Mol. Cell Biol. 10:4384-4388.

93. Ruezinsky, D., Beckmann, H. and Kadesch, T. (1990) Genes Dev. 5:29-37.

94. Kadesch, T. (1992) Immunol. Today 13:31-36.

95. Pongubala, J.M.R. and Atchison, M.L. (1991) Mol. Cell Biol. 11:1040-1047.

96. Nelson, K.J., Kelley, D.E. and Perry, R.P. (1985) Proc. Natl. Acad. Sci. USA 82:5305-5309.

97. Baeuerle, P.A. and Baltimore, D. (1989) Genes Dev. 3:1689-1698.

98. Kawakami, K.,Scheidereit, C. and Roeder, R.G. (1988) Proc. Natl. Acad. Sci. USA 85:4700-4704.

99. Lenardo, M.J., Kuang, A.,Gifford, A. and Baltimore, D. (1988) Proc. Natl. Acad. Sci. USA 85:8825-8829.

100. Baeuerle, P.A. and Baltimore, D. (1988) Science 242:540-546.

101. Ghosh, S. and Baltimore, D. (1989) Science 344:678-682.

102. Lenardo, M.J. and Baltimore, D. (1989) Cell 58:227-229.

103. Lenardo, M.J., Pierce, J.W., and Baltimore, D. (1987) Science 236:1573-1577.

104. Pettersson, S., Cook, G.P., Brugemann, M., Willians, G.T., Neuberger, M.S. (1990) Nature 344:165-168.

105. Meyer, K.B. and Neuberger, M.S., (1989) EMBO J. 8:1959-1964.

106. Luria, S., Gross G., Horowitz, M. and Givol, D. (1987) EMBO J. 6:3307-3312.

107. Ho, I.-C., Yang, L-H., Morle, G. and Leiden, J.M. (1989) Proc. Natl. Acad. Sci. USA 86:6714-6718.

108. Winoto, A. and Baltimore, D. (1989) EMBO J. 8:729-733.

109. Ho, I-C. and Leiden, J.M. (1990) Mol. Cell Biol. 10:4720-4727.

110. Leiden, J.M. (1992) Immunol. Today 13:22-30.

111. Gottschalk, L.R. and Leiden, J.M. (1990) Mol. Cell Biol. 10:5486-5495.

112. Van de Wetering, M., Oosterwegel, M., Dooijes, D. and Clevers, H. (1991) Genes Dev. 5:880-894.

113. Waterman, M.L., Fischer, W.H., and Jones, K.A. (1991) Genes Dev. 5:1-4.

114. Travis, A., Amsterdam, A., Belanger, C., and Grosschedl, R. (1991) Genes Dev. 5:880-894.

115. Ho, I-C, Bhat, N.K., Gotlschal, L.R. et al. (1990) Science 250:814-818.

116. Bhat, N.K., Fisher, R.J., Fujiwara, S., Ascione, R. and Papas, T.S. (1987) Proc. Natl. Acad. Sci. USA 84:3161-3165.

117. Karim, F.D.,Urners, L.D., Thummer, C.S., et al. (1990) Genes Dev. 4:1451-1453.

118. Ho, I-C., Voorhees, P., Marin, N., et al. (1991) EMBO J. 10:1187-1192.

119. Yamamoto, M., Ko, L.J., Leonard, M.W., et al. (1990) Genes Dev. 4:1650-1662.

120. Ko, L.J., Yamamoto, M., Leonard, M.W., et al. (1991) Mol. Cell Biol. 11:2778-2784.

121. Winoto, A. and Baltimore, D. (1989) Cell 59:649-655.

122. Dimond, D.J., Nelson, F.B. and Reinherz, E.L. (1989) J. Exp. Med. 169:1213-1231.

123. Anderson, S.J., Chou, H.S. and Loh, D.Y. (1988) Proc. Natl. Acad. Sci. USA 85:3551-3554.

124. Anderson, S.J., Miyaki, S. and Loh, D.Y. (1989) Mol. Cell Biol. 9:4835-4845.

125. Royer, H.D. and Reinherz, E.L. (1987) Proc. Natl. Acad. Sci. USA 84:232-236.

126. Krimpenfort, P., de Jong, R., Uematsu, Y., et al. (1988) EMBO J. 7:745-750.

127. McDougall, S., Peterson, C.L. and Calame, K. (1988) Sciencs 241:205-208.

128. Gottschalk, L.R. and Leiden, J.M. (1990) Mol. Cell. Biol. 10:5486-5495.

129. Hoshimoto, Y. (1989) J. Exp. Med. 169:2097-2107.

130. Tonegawa, S., Berns, A., Bonneville, M., et al. (1989) Cold Spring Harb. Symp. Quant. Biol. 54:31-44.

131. Kappes, D.J., Browne, C.P. and Tonegawa, S. (1991) Proc. Natl. Acad. Sci. USA 88:2204-2208.

132. Bories, J.C., Loiseau, P., d'Auriol, L., et al. (1990) J. Exp. Med. 171:75-83.

133. Redondo, J.M., Hata, S., Brocklehurst, C. and Krangel, M.S. (1990) Science 247:1225-1229.

134. Hunkapiller, T., Goverman, J., Koop, B.F. and Hood, L. (1989) Cold Spring Harb. Symp. Quant. Biol. 54:15-29.

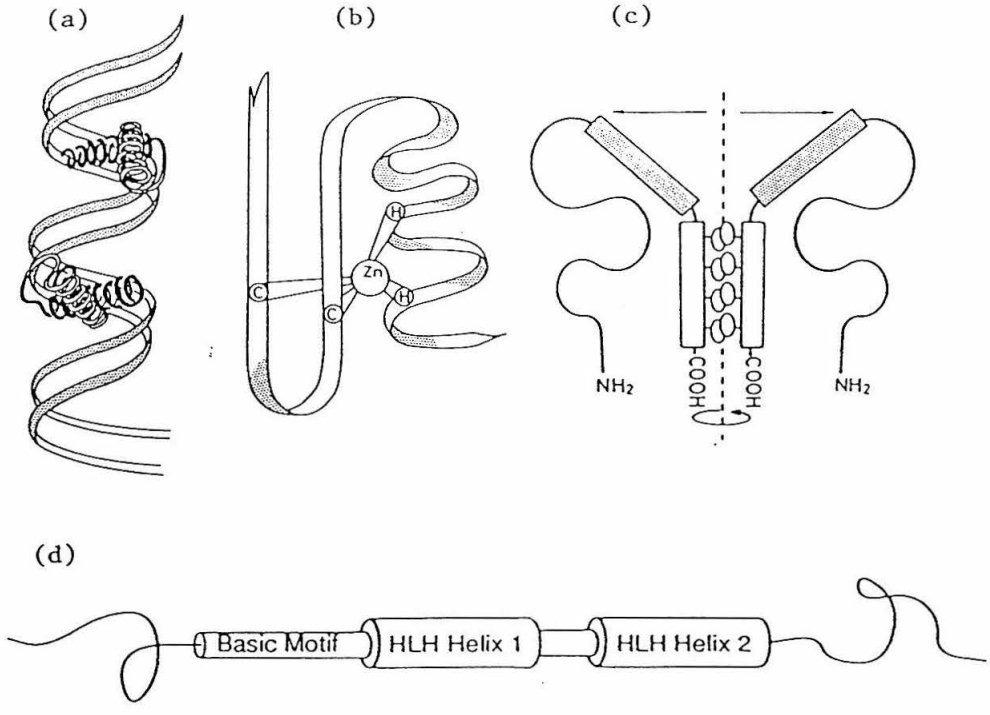
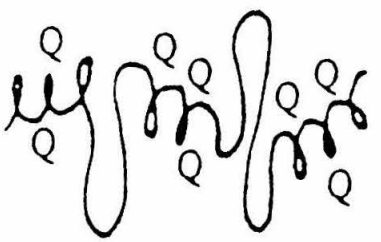

(a)
(b)
(c)
Zn
H
H
C
C
NH2
COOH
COOH
NH2
(d)
Basic Motif
HLH Helix 1
HLH Helix 2

*Figure 1* Four types of DNA binding motifs (adapted from Ref. 7). (a) A helix-turn-helix protein folds in the major groove of the DNA double helix, where it recognizes half of the dyad-symmetric binding site. Dimerization of helix-turn-helix protein is necessary for stable interaction with DNA sequence. (b) A zinc finger binds to a zinc ion though paired cysteine and histidine residues. The cysteins are located in antiparallel β sheets. The histidine residues are in an α-helical structure. (c) Two polypeptides of a leucine zipper protein dimerize via hydrophobic interactions between leucine residues (open ovals) located in α helices. The dark rectangles represent the basic regions that recognize the DNA sequence. (d) Helix-loop-helix (HLH) protein has two amphipathic α helices that are separated by a loop. A basic region that is adjacent to the HLH structure is necessary for DNA binding.

(a)

(b)




(c)

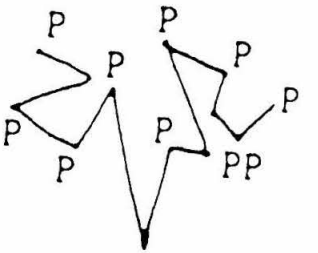


*Figure 2* Three types of transcription activation domains of transcription factors (adapted from Ref. 43). (a) Acidic domain that is rich in negatively charged aspartic acids and glutamic acids. (b) Glutamine-rich domain, glutamine (Q) is indicated. (c) Proline-rich domain, "P" represents proline.

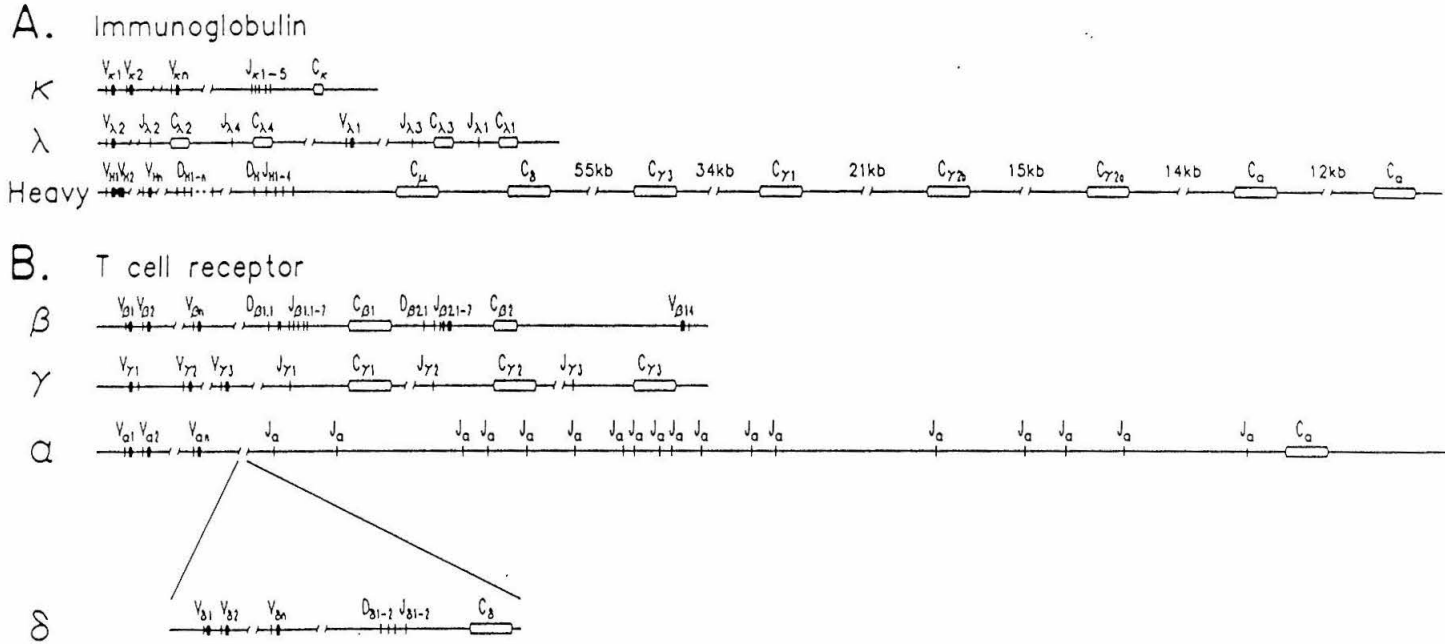

A.
Immunoglobulin
κ
λ
Heavy
55kb
34kb
21kb
15kb
14kb
12kb
B.
T cell receptor
β
γ
α
δ

*Figure 3* Organization of the murine immunoglobulin (A) and T cell receptor (B) gene families (adapted from Ref. 134). Filled boxes are variable region gene segments, and open boxes are constant region genes. Numbers are distances in kilobase. Double slashes without numbers indicate unknown distances. The TCR δ gene family is shown as an insert to the TCR α gene family.

Table 1 DNA recognition sequences for Ig transcription factors

| Motif | Location | Transcription factor family |
|---|---|---|
| **Octamer** | | **POU-domain** |
| ATTTGCAT | Ig promoter | OCT-1/OCT-2/OCT-3 |
| ATTTGCAT | IgH enhancer | OCT-1/OCT-2/OCT-3 |
| ATTGACAT | Igκ enhancer | OCT-1/OCT-2/OCT-3 |
| **κB** | | **REL-related** |
| GGGGACTTTC | Igκ enhancer | NF-κB |
| **E-box** | | **Helix-loop-helix** |
| CAGGTGTT (μE5) | IgH enhancer | E12/E47/ITF-1/ITF-2 |
| CAGGTGGC (κE2) | Igκ enhancer | E12/E47/ITF-1/ITF-2 |
| CATGTGGC (μE3) | IgH enhancer | TEF3/TFEB/USF |
| CATGTGGC (κE3) | Igκ enhancer | TEF3/TFEB/USF |

Table 2 DNA recognition sequences for TCR transcription factors

| Motif | Location | Transcription factor family |
|---|---|---|
| **CRE** | | **Leucine zipper** |
| TGACGTCA | TCR α enhancer Tα1 | CREB/CRE-BPI/ATF-2/mXBP/ATF-4 |
| TCACATCA | TCR β enhancer Tβ2 | CREB/CRE-BPI/ATF-2/mXBP/ATF-4 |
| TG<br>AGTGA TCA<br>CA | TCR β promoter | CREB/CRE-BPI/ATF-2/mXBP/ATF-4 |
| **TCF-1** | | **HMG box** |
| CCCTTTGAAG | TCR α enhancer Tα2 | TCF-1/TCF-1α/LEF-1 |
| ACTTCAGAGGG | TCR β enhancer Tβ5 | TCF-1/TCF-1α/LEF-1 |
| CCCTTTGA | TCR δ enhancer δE7 | TCF-1/TCF-1α/LEF-1 |
| **Ets** | | **Basic domain/α helix** |
| CATCCTCT | TCR α enhancer Tα2 | Ets-1 |
| AACAGGATGT | TCR β enhancer Tβ3 | Ets-1 |
| **GATA** | | **Zinc finger** |
| TGATAG | TCR α enhancer Tα3 | GATA-3 |
| TTATCTC | TCR β enhancer Tβ2 | GATA-3 |
| TGATAA | TCR δ enhancer δE4 | GATA-3 |

# Chapter 1

Characterization of a cDNA Encoding a Zinc Finger Protein That Binds to the Human T Cell Receptor Gene V$\beta$8.1 Promoter

**ABSTRACT**

A cDNA clone ht$\beta$, encoding a zinc finger protein that binds to the promoter region of the human T cell receptor (TCR) gene V$\beta$8.1, was cloned from a human peripheral blood T cell library. The fragment containing four zinc fingers of the class $Cys_2$-$X_{12}$-$His_2$ may be responsible for DNA binding to the TCR V$\beta$8.1 promoter sequence GAAGTTGGGGGTGGTG. A putative transcriptional activation domain that is highly negative charged also was found in ht$\beta$. Analysis of expression of ht$\beta$ mRNA revealed similar expression levels in Hela cells, Jurkat T cells, Ramos B cells and the U-937 monocyte line. Interestingly, ht$\beta$ protein also binds to other T cell receptor gene regulatory sequences, implying that ht$\beta$ may have multiple functions on TCR gene expression.

## INTRODUCTION

T cells recognize antigens in the context of self major histocompatibility complex (MHC) using a clonally expressed T cell receptor (TCR). The genes encoding TCR-$\alpha$, -$\beta$, -$\gamma$ and -$\delta$ chains have been characterized (1-10). The four TCR genes are similar to immunoglobulin (Ig) genes in that each consists of multiple germ line gene segments that rearrange during T cell development to generate a mature T cell population with clonally distributed receptors (11). Although both T and B cells derive from common bone marrow precursors, the expression of Ig and TCR genes is usually lineage specific, for example, T cells rearrange and express TCR but not Ig genes. Also, previous studies have suggested that both T and B cells use a common mechanism for the assembly of TCR and Ig genes (12).

Every TCR gene has two parts, a variable gene region (V) and a constant gene region (C). The variable gene is generated through a DNA rearrangement event. The V $\alpha$ and V$\gamma$ genes are produced by combining a V gene segment and a joining (J) gene segment, then the V$\alpha$J$\alpha$ or V$\gamma$J$\gamma$ joins the C gene upon mRNA splicing. V$\beta$ and V$\delta$ genes are constructed from three gene segments, a V, a diversity (D) and a J gene segment. The D and J gene segments join first, followed by V and DJ joining. Finally, the V$\beta$D$\beta$J$\beta$ or V$\delta$D$\delta$J$\delta$ combines with the C gene upon mRNA splicing. The TCR $\delta$ genes are located within the TCR $\alpha$ locus, and TCR $\beta$ and $\gamma$ genes are mapped on different chromosomes (11,13). TCR gene rearrangement and expression can be divided into four stages (13). In fetal thymus, TCR $\delta$ and $\gamma$ gene segments rearrange first. Those T cells that succeed in productively rearranging $\delta$ and $\gamma$ genes become $\gamma\delta$ T cells. Those that fail to produce $\gamma$ and $\delta$ chains may then produce a D$\beta$-J$\beta$ joining, followed by rearrangement of a V$\beta$ gene segment to D$\beta$-J$\beta$ to form a variable region gene,V$\beta$-D $\beta$-J$\beta$. Finally the V$\alpha$ and J$\alpha$ gene segments rearrange to produce V$\alpha$-J$\alpha$. Those T cells that succeed in rearranging TCR $\beta$ and $\alpha$ loci express an $\alpha\beta$ heterodimer on their

surface, becoming $\alpha\beta$ T cells. While the organization of four TCR genes has been well established, the regulation of rearrangement and expression of these genes in the different T cell subsets is relatively unknown. Recent studies have described transcriptional enhancers and silencers in the human and murine TCR genes (14-18), and two DNA binding proteins that interact with the human TCR $\alpha$ gene enhancer (19-20). Several studies have addressed the question of tissue specificity of V$\beta$ promoters. One study reveals that the human TCR V$\beta$8.1 promoter is only active in T cells (21), implying that there are some T cell specific transcription factors that bind to the identified V$\beta$8.1 promoter. A conserved decamer motif 10-40 bp upstream of the TATA box has been identified in the murine V$\beta$ promoters (22). This decamer also is present in the human V$\beta$8.1 promoter (23). Royer and Reinherz have shown that there are four regions to which transcription factors bind in the 175 bp fragment of the human V$\beta$8.1 promoter. One of the regions, a 21 bp GT-rich motif 72 bp upstream of the transcription initiation site was only protected by nuclear extracts from T cell lines, not other cells in a DNAse footprinting study (24). The result suggests there may be T cell specific transcription factors binding to the 21 bp GT-rich motif.

We now report the isolation and characterization of a cDNA encoding a DNA binding protein that recognizes the 21 bp GT rich motif in the promoter region of the human TCR V$\beta$8.1 gene. This human TCR $\beta$ gene promoter binding protein (ht$\beta$) contains four Cys-His zinc fingers.

## MATERIALS AND METHODS

### *Library screening*

A $\lambda$gt11 cDNA library prepared with mRNA from PHA-stimulated human peripheral blood T cells (Clontech) was screened according to the method of Vinson et

al. (25). Plaques ($2x10^6$) were screened at a density of $2x10^4$ per filter after lytic infection of *E. coli* strain Y1090. The probe for screening the λgt11 cDNA library was generated starting with the two complementary oligonucleotides corresponding to the sequence of human TCR Vβ8.1 promoter element -72 to -92 (24):

```
5' AATTTTAAAGAAGTTGGGGGTGGTG 3'
   3' AATTTCTTCAACCCCCACCACTTAA 5'
```

The probes used in the screening were prepared by nick translation of size-selected, double-stranded catenated oligonucleotides using [$\alpha^{32}$P] dATP and [$\alpha^{32}$P] dCTP (26). The EcoRI inserts from positively selected plaques were subcloned into Bluescript-KS II (-) for sequence analysis using the dideoxy chain termination method for both strands (26). A λgt11 cDNA library made with Jurkat T cell line derived mRNA following the procedure of cDNA SYNTHESIS SYSTEM PLUS (Amersham) was constructed for isolating full length htβ cDNA.

*Southern and Northern analysis*

Total genomic DNA was isolated from the human T cell lymphoma-derived cell line Jurkat. DNA (10 μg) was digested by EcoRI, HindIII, BamHI and XbaI, resolved on a 0.8% agarose gel, and transferred to nylon membrane (Zeta-Probe, Bio-Rad). Southern hybridization was carried out at 37°C overnight with $2.5x10^6$ cpm/ml denatured probe (5xSSC, 20 mM Sodium Phosphate pH 6.7, 10% Dextran Sulfate, 1% SDS, 0.5% powdered milk, 100 μg/ml denatured salmon sperm DNA, and 50% formamide). The washes were carried out in 0.1xSSC, 0.1% SDS at 65°C for 15 minutes, and repeated twice.

The sources of the human RNAs were T cell line Jurkat (ATCC CRL 8163), B cell line Ramos (ATCC CRL 1596), Monocyte line U-937 (ATCC CRL 1593), and *Hela* cells (ATCC CCL2). RNA was isolated from these cell lines by the guanidinium method (26). and poly $(A)^+$ RNAs were selected by oligo(dT)-cellulose chromatography (27).

For Northern blot analysis, 10$\mu$g of poly (A)$^+$ RNA was resolved on a formaldehyde gel (26) and transferred to nylon membrane (Zeta-probe GT, Bio-Rad). The hybridization and washes were performed following the procedure recommended by the membrane manufacturer (Bio-Rad). RNA molecular weight markers were purchased from GIBCO BRL.

*Bacterial expression*

The bacteriophage T7 expression system was used for production of the ht$\beta$ protein (28). A DNA fragment containing a portion of the ht$\beta$ coding sequence, including the four zinc fingers, was created with NdeI and BclI sites at the ends using appropriate oligonucleotides and PCR amplification. This was cloned into the NdeI and BamH I sites of the vector pET3a. This plasmid was transformed into the strain BL21 (DE3) that has the T7 RNA polymerase gene under *lacuv5* promoter control. Transformed bacteria were induced with 1 mM IPTG at OD600=0.7 for 4 hrs and harvested. Extracts were prepared according to the method of Hoey and Levine (29), modified by Treisman and Desplan (30).

*DNase I footprinting*

The probes for DNase I footprinting were either RsaI-XbaI or HinfI-NcoI fragments from the human V$\beta$8.1 promoter region. They were isolated from cosmid H7.1 (23) and labeled at one end using the filling-in method with Klenow DNA polymerase (26). DNase I footprinting assays were performed as described by Galas and Schmitz (31). The binding reactions were done with 3-5 ng of $^{32}$P-labeled DNA and 2 $\mu$g of poly(I-C) in 30 $\mu$l for 20 minutes on ice. The composition of the binding reaction was according to Fan and Maniatis (32). After binding, 4 $\mu$l of 100 mM $MgCl_2$ and 150mM $CaCl_2$ were added to the reaction, followed by 1 $\mu$l of freshly diluted DNase I (Worthington) at a final concentration of 25 $\mu$g/ml. DNase I digestion was stopped after

1 minute on ice by addition of stop solution described by Hoey and Levine (29). The samples were extracted twice with phenol-chloroform (1:1), chloroform once, ethanol precipitated and electrophoresed in a 6% polyacrylamide/8M urea gel.

## RESULTS

### *Isolation of the htβ gene*

The DNA probe containing the sequence -72 to -92 of the human Vβ8.1 promoter was used to screen a λgt11 expression library constructed from T cell derived mRNA. The screening of $2.6 \times 10^6$ plaques yielded two clones. One of the clones, htβ16 contained a 1.8 kb insert which was subcloned into BS-KS II (-) for sequence analysis. Using the 1.8 kb htβ16 insert to rescreen the same library, we found a clone htβ8 which overlaps with the 3' sequence of htβ16. Another clone htβ3, which has a 1.8 kb insert, overlaps for 1 kb with the 5' sequence of htβ16 (fig. 1). It was isolated from a λ gt11 library made with Jurkat T cell line derived mRNA. Southern blotting of Jurkat cell line genomic DNA digested by EcoRI, BamHI, HindIII and XbaI and probed with the htβ16 1.8 kb fragment indicates a dominant single band, although some faint bands appear upon longer exposure (fig. 2).

### *DNA sequence of the htβ gene*

The DNA sequence of htβ was determined (fig. 3). The sequence predicts a long open reading frame of 1362 nucleotides that is terminated by six in-frame stop codons. The 3' untranslated region is 666 nucleotides long . The poly A site was not identified. The nucleotide sequence of the open reading frame was translated into its corresponding 454 amino acid sequence. The molecular weight of the htβ protein encoded by the full length mRNA is approximately 49 kd. Inspection of the amino acid sequence shown in

figure 3 indicates that the ht$\beta$ protein has four tandem zinc fingers near the middle of the open reading frame. The first two finger motifs have the general form of Cys-$X_2$-Cys-$X_3$-Phe-$X_5$-Leu-$X_2$-His-$X_3$-His, originally described in the RNA polymerase III transcription factor TFIIIA (33), and also found in several DNA-binding protein domains (34). The third and fourth fingers have the modified form of Cys-$X_2$-Cys-$X_8$-His-$X_3$-His and Cys-$X_2$-Cys-$X_3$-Phe-$X_6$-Leu-$X_1$-His-$X_4$-His respectively (fig.4). At the amino terminal side of the zinc finger region there is a highly acidic segment (residues 54-99) that has potential of forming an $\alpha$-helix. This region has a net negative charge of -12. Other features of interest in the ht$\beta$ sequence include a highly basic segment (residues 129-153) at the amino terminal side of the zinc finger region and another basic segment (residues 314-344) downstream of the fingers.

*Ht$\beta$ encodes a protein that can bind to the human TCR V$\beta$8.1 promoter*

A fragment of ht$\beta$ (residues 81-445) containing most of the coding region was expressed in *E. coli*. A portion of the acidic region is missing from the expressed protein but all four zinc fingers are present. The plasmid pET3a-ht$\beta$ encodes a 40 kd protein composed of 365 amino acids. The deduced molecular weight of this protein is the same as the observed size from SDS-polyacrylamide gel analysis (fig. 5).

Crude lysates of E. coli strain BL21 (DE3) containing pET3a-ht$\beta$ were tested for DNA binding activity by DNAse I footprinting with the human TCR V$\beta$8.1 promoter element (fig. 6). The extracts prepared from BL21 (DE3) (pET3a-ht$\beta$), but not BL21 (DE3) (pET3a) were found to protect the 16 nucleotides GAAGTTGGGGGTGGTG of the V$\beta$8.1 promoter from digestion by DNase I. The DNase I footprinting assay was done with both coding and noncoding strands of the V$\beta$8.1 promoter, and the protected regions were about the same. The sequence of the ht$\beta$ binding site detected by the footprinting assay is identical to that of the probe used for screening the $\lambda$gt11 expression library. Also, it appears that the 365 amino acid residues of the truncated ht

$\beta$ protein that includes the four zinc fingers are sufficient for the sequence-specific binding of ht$\beta$ protein to DNA.

*Expression of ht$\beta$ mRNA*

To study the expression pattern of the ht$\beta$ gene, RNA was isolated from various tissue culture cell lines. Poly $(A)^+$ RNAs were selected by oligo (dT)-cellulose chromatography and detected by Northern blot analysis using the 1.8 kb segment of ht$\beta$ 16 as a probe. As shown in figure 7, three RNA species were detected, with estimated sizes of 4.2, 7.6 and 8.6 kb. The same blot was rehybridized with the EcoRI-EcoRV fragment of ht$\beta$3 that does not contain the zinc finger region, and again three RNA species of the same size as in figure 7 were seen. To control for potential variation on the level of RNA electrophoresed, the blot was rehybridized with a human $\beta$-actin probe. Comparable levels of hybridized RNA were observed with the actin probe indicating the expression level of ht$\beta$ is not significantly different in these four cell lines.

## DISCUSSION

An oligonucleotide probe to the sequence of the human TCR V$\beta$8.1 promoter element -72 to -92 was used to isolate a cDNA clone from a $\lambda$gt11 expression library prepared from human peripheral T cell mRNA. This cDNA hybridizes to three mRNA species present in a variety of cell types, including Hela cells, U-937 monocyte line, Ramos B cell line and Jurkat T cell line. The three mRNA species may reflect alternative mRNA splicing of the ht$\beta$ gene or the 7.6 and 8.6 kb mRNA species could be unspliced forms of the 4.2 kb RNA. Other cDNA segments containing overlapping sequences were isolated from the above library or a library prepared from Jurkat cell

line RNA. The sequence of a composite cDNA of 3.6 kb was partially determined. Translation of this sequence reveals a long open reading frame of 454 amino acid residues encoding a protein of approximately 49 kd. Ht$\beta$ may be a single copy gene since Southern blot analysis using digests of human genomic DNA revealed only one band. Longer film exposure reveals more bands for each digestion. The ht$\beta$ probe may cross hybridize to other genes encoding zinc finger proteins in the genome.

The DNA binding domain of the ht$\beta$ protein has been localized to a 118 amino acid protein fragment that contains four zinc fingers of $Cys_2$-$X_{12}$-$His_2$. This conserved motif, found in many DNA binding proteins, was first described in transcription factor TFIIIA (33). The consensus sequence of the four fingers is Cys-$X_2$-Cys-$X_3$-Phe-$X_5$-Leu-$X_2$-His-$X_3$-His. Although crystal structure analysis of a zinc finger-DNA complex suggests that one finger contacts 3 bp (36), the ht$\beta$ protein recognizes a sequence GAAGTTGGGGGTGGTG of approximately 4 bp per finger domain, which also is different from analysis of TFIIIA binding to the 5S gene where each domain recognizes a 5.5 bp region (33). Examination of the sequences on the ht$\beta$ fingers and their binding sites suggests arginine-guanine contacts also may be important for forming the finger and DNA complex, as noted in the crystal structure of the mouse immediate early protein Zif268 (36). The sequence analysis of the ht$\beta$ protein suggests the negatively charged segment that can form an $\alpha$-helical structure may be involved in transcriptional activation as described in the yeast factors GCN4 and GAL4 (37-39), AP-1/Jun transcription factors (40-41) and immunoglobulin heavy chain enhancer binding protein (42). The two basic segments may be involved in nuclear localization of ht$\beta$ as proposed by Dingwall and Laskey (43).

The specific binding of ht$\beta$ protein to the human V$\beta$8.1 promoter was shown with a DNase I footprinting assay (fig. 5). Since the portion of ht$\beta$ protein that includes the zinc finger region is able to bind DNA specifically, the finger region is most likely responsible for DNA binding. The binding site is approximately 16 nucleotides,

GAAGTTGGGGGTGGTG, which is a little different from the site of 21 nucleotides, TTAAAGAAGTTGGGGGTGGTG that was detected with T cell line nuclear extracts (24). This may be explained by other factors besides htβ protein in the T cell nuclear extracts that bind to the sequence around htβ protein binding sites. Although we do not have data on whether htβ protein binds as a dimer to DNA, the lack of symmetrical structure at the binding site makes this unlikely.

The data presently available on the htβ gene suggest that it may encode a protein with transcriptional regulation function, but how it regulates TCR gene expression remains to be determined. In addition to the human TCR Vβ8.1 promoter, the htβ binding site is also present in the human Vβ8.2 gene promoter. Whether htβ protein can bind to other human Vβ promoters needs to be tested. Interestingly, htβ protein can bind to the mouse TCR α silencer characterized by Winoto et al. (16) (data not shown). This observation indicates that htβ protein may have multiple functions in regulating TCR gene expression. Royer et al. previously reported a nuclear extract activity only from T cell lines and not other cell lines that could bind the same region as the htβ protein binding site also determined by DNase I footprinting analysis (24). We detected htβ mRNA in T cells, B cells and other cell lines.

The question of whether the htβ protein plays a role as a T cell specific or general transcription factor is unresolved. Although htβ mRNA is expressed in various cell lines, there may be different htβ protein products in different cell lines due to post-translational modifications. Htβ gene products may have functions in T cells that are different from those in other cells. It also is possible that htβ protein functions as a general transcription factor. Since the htβ protein binding site is rich in GT, it may bind to other GT-rich sequences (GT boxes), just as Sp1 binds to GC boxes (43). If this is the case, htβ protein may have more functions in regulating transcription. Finally, the isolation of the htβ gene provides us with another system to study zinc finger and DNA interactions. As more genes encoding zinc finger proteins are cloned, how each

finger specifically contacts certain nucleotides can be determined.

REFERENCES

1. Yanagi, Y., Yoshikai, Y., Leggette, K., Clark, S., Aieksander, I. and Mak, T.W. (1984) Nature 308:145-149.

2. Hedrick, S.M., Nielsen, E.A., Kavaler, J., Cohen, D.I. and Davis, M.M. (1984) Nature 308:149-153.

3. Chien, Y.H., Becker, D.M., Lindsten, T., Okamura, M. Cohen, D.I. and Davis, M.M. (1984) Nature 312:31-35.

4. Saito, H.D., Kranz, M., Takagaki, Y., Hayday, A.C., Eisen, H.N. and Tonegawa, S. (1984) Nature 312:36-40.

5. Sim, G.K., Yague, J., Nelson, J., Marrack, P., Palmer, E., Augustin, A. and Kappler, J. (1984) Nature 312:771-75.

6. Saito, H.D., Kranz, M., Takagaki, Y., Hayday, A.C., Eisen, H.N. and Tonegawa, S. (1984) Nature 309:757-762.

7. Kranz, D.M., Satio, H., Heller, M., Takagaki, Y., Haas, W., Eisen, H.N. and Tonegawa, S. (1985) Nature 313:752-755.

8. Chien, Y.H., Iwashima, M., Kaplan, K.B., Elliott, J.F. and Davis, M.M. (1987) Nature 327:677-682.

9. Hata, S., Brenner, M.B. and Krangel, M.S. (1987) Science 238:678-682.

10. Chien, Y.H., Iwashima, M., Wettstien, D.A., Kaplan, K.B., Elliott, J.F., Born, W. and Davis, M.M. (1987) Nature 330:722-727.

11. Kronenberg, M., Siu, G., Hood, L. and Shastri, N. (1986) Annu. Rev. Immunol. 4:529-591.

12. Yancopoulus, G.D., Blackwell, T.K., Suh, H., Hood, L. and Alt, F.W. (1986) Cell 44:251-259.

13. Davis, M.M. and Bjorkman, P.J. (1988) Nature 334:395-402.

14. Ho, I-C., Yang, L-H., Morle, G. and Leiden, J.M. (1989) Proc. Natl. Acad. Sci. USA

86:6714-6718.

15. Winoto, A. and Baltimore, D. (1989) EMBO J. 8:729-733.

16. Winoto, A. and Baltimore, D. (1989) Cell 59:649-655.

17. Krimpenfort, P., de Jong, R., Uematsu, Y., Dembic, Z., Ryer, S., von Boehmer, H., Steinmetz, M. and Berns, B. (1988) EMBO J. 7:745-750.

18. McDougall, S., Peterson, C.L. and Calame, K. (1988) Science 241:205-208.

19. Waterman, M.L. and Jones, K.A. The New Biologist (1990) 2:621-636.

20. Ho, I-C., Bhat, N.K., Gottschalk, L.R., Lindsten, T., Thompson, C.B., Papas, T.K. and Leiden,J.M. Science (1990) 250:814-818.

21. Diamond, D.J., Nelson, P.B. and Reinherz, E.L. (1989) J. Exp. Med. 169:1213-1231.

22. Anderson, S.J., Chou, H.S. and Loh, D.Y. (1988) Proc. Natl. Acad. Sci. USA 85:3551-3554.

23. Siu, G., Strauss, E.C., Lai, L. and Hood, L. (1986) J. Exp. Med. 164:1600-1614.

24. Royer, H.D. and Reinherz, E.L. (1987) Proc. Natl. Acad. Sci. USA 85:232-236.

25. Vinson, C.R., Lamarco, K.L., Johnson, P.E., Landschulz, W.H. and McKnight, S.L. (1988) Genes Dev. 2:801-806.

26. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1989) Molecular Cloning, A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

27. Aviv, H. and Leder, P. (1972) Proc. Natl. Acad. Sci. USA 69:1408.

28. Studier, F.W., Rosenberg, A.H., Dunn, J.J., and Dubendorff, J.W. (1990) Methods Enzymol. 185:60-89.

29. Hoey, T. and Levine, M. (1988) Nature 332:858-861.

30. Treisman, J. and Desplan, C. (1988) Nature 341:335-337.

31. Galas, D.J. and Schmitz, A. (1978) Nucl. Acids Res. 5:3157-3171.

32. Fan, C.-M., and Maniatis T. (1990) Genes Dev. 4:29-42.

33. Miller, J., McLachlan, A.D. and Klug, A. (1985) EMBO J. 4:1609-1614.

34. Evans, R.M. and Hollenberg, S.M. (1988) Cell 52:1-3.

35. Gaul, U., Seifert, E., Schun, R. and Jackle, H. (1987) Cell 50:639-647.

36. Pavletich, N.P. and Pabo, C.O. (1991) Science 252:809-817.

37. Hope, I.A. and Struhl, K. (1985) Cell 43:177-188.

38. Ma, J. and Ptashne, M. (1987) Cell 51:113-119.

39. Ptashne, M. (1988) Nature 335:683-689.

40. Bohmann, D., Bos, T.J., Admon, A., Nishmura, T., Vogt, P.K., and Tjian, R. (1987) Science 238:1386-1392.

41. Struhl, K. (1988) 332:649-650.

42. Beckmann, H., Su, L.-K. and Kadesch, T. (1990) Genes Dev. 4:167-179.

43. Dingwall, C. and Laskey, R.A. (1986) Ann. Rev. Cell Biol. 2:367-390.

44. Kadonaga, J.T., Carner, K.R., Masiarz, F.R. and Tjian, R. (1987) Cell:51:1079-1090.

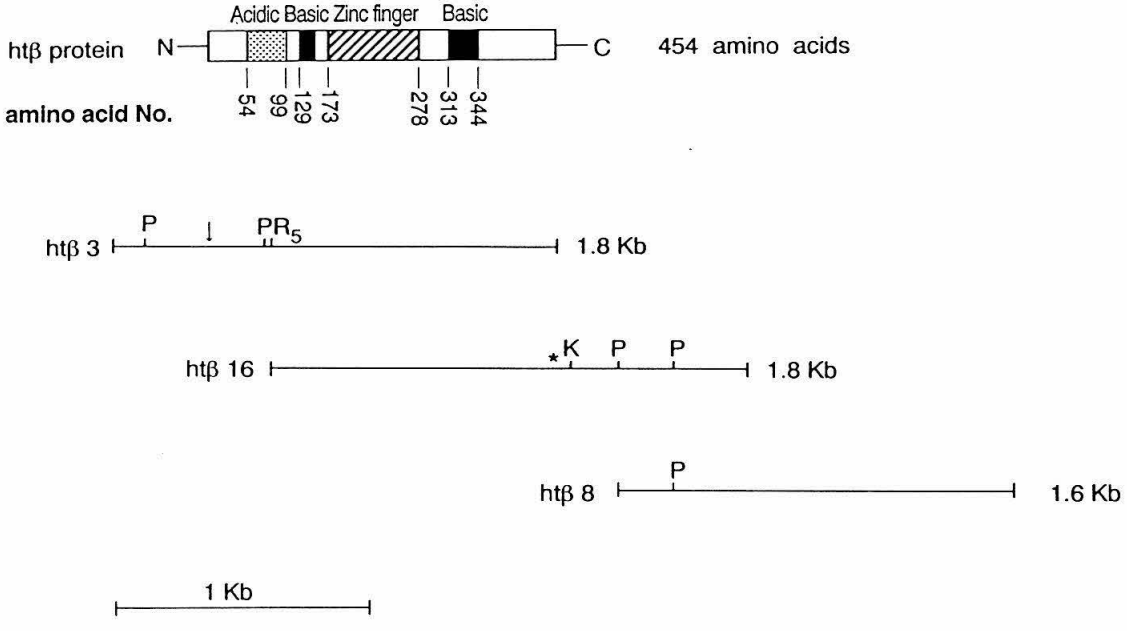

Acidic Basic Zinc finger Basic
htβ protein
N
C
454 amino acids
amino acid No.
54
99
129
173
278
313
344
P
PR5
htβ 3
1.8 Kb
K
P
P
htβ 16
1.8 Kb
P
htβ 8
1.6 Kb
1 Kb

*Figure 1* cDNAs and predicted protein structure of htβ. The overlapping cDNAs htβ16 and htβ8 were isolated from a human peripheral T cell cDNA library, and cDNA htβ3 was isolated from a Jurkat cell line cDNA library. Restriction enzyme sites: K, KpnI; R1, EcoRI; R5, EcoRV; P, PstI. A schematic diagram of the predicted htβ protein structure is shown, with three cDNA clones aligned to the protein structure. The interesting motifs in the protein are indicated. "↓" indicates the translation start site and "*" indicates the stop codon.

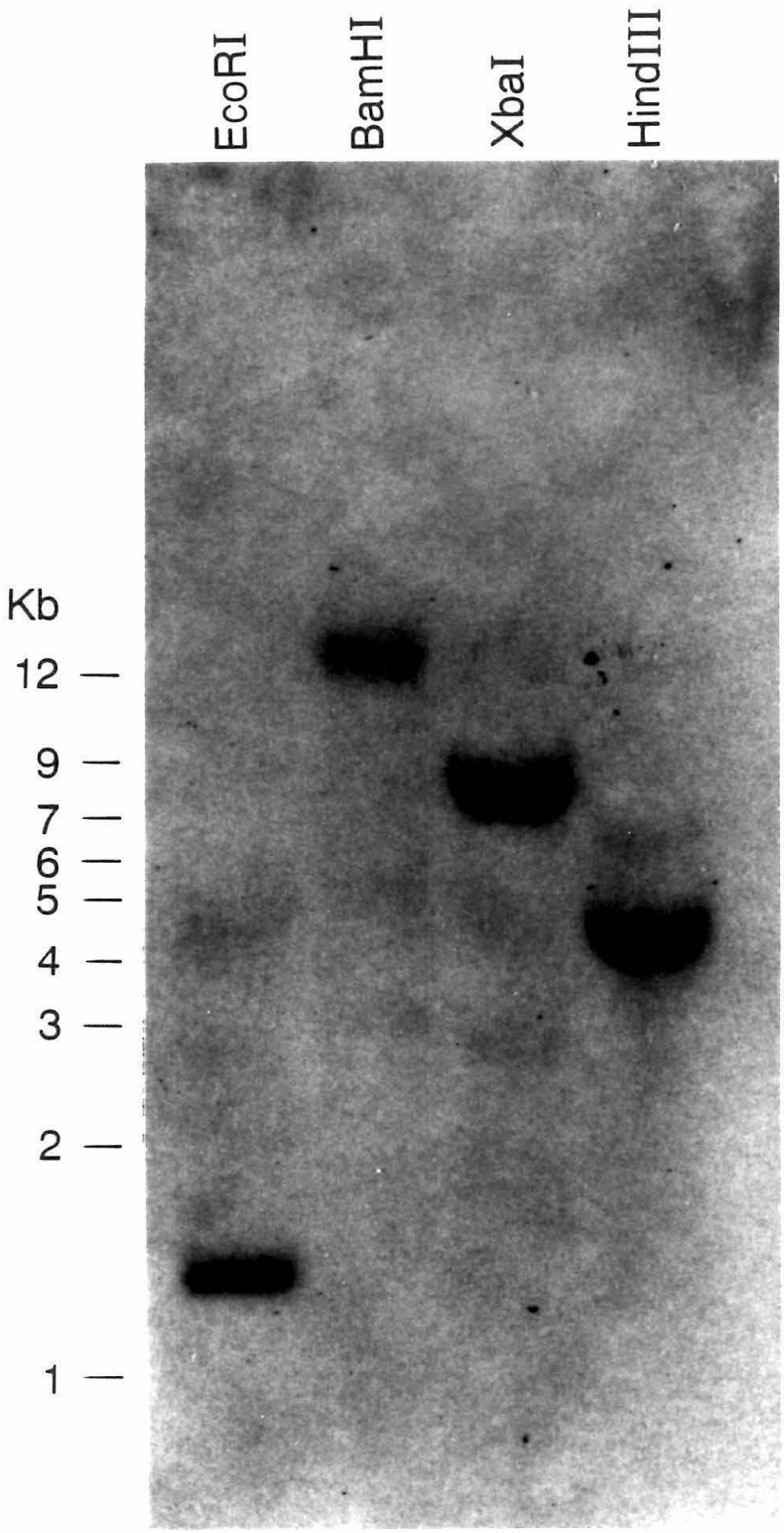

EcoRI
BamHI
XbaI
HindIII
Kb
12
9
7
6
5
4
3
2
1

*Figure 2* Southern blot analysis of Jurkat genomic DNA probed with the htβ 16 clone.

```
1     GAATTCCGGAGAAAGGCGCAGGGGTGGGAGCTGTTGCCGAAGCTGCCACAGCAAAAGTTC   60

61    TCCCCCCTCCCCCCTTCCCCTCCTCTCAAGGCCCCTAGAAAGGTTGGAGCTGCCGGCCCT  120

121   GCAGTCGGTGACCGCTGACGACTTCGGCCGCGCCCGCGGATAGAGGGAGGAATCAGCAGC  180

181   TTGGAAATTCAAGCACGTGATCTGGCGGGATGGCGTTTGCCTAACGTATTTAATGGAGGA  240

241   ATCGGATGGCATAAGTGATTAAGGTGGTATTGAGGATTTCTGAAGCCTATGAAAGGTAGA  300

301   AACTCAACCATGATTTCTTTTTCAACTCTACAGCATTCCTTTCCTTGAAGTCTTCGTTTT  360

361   TACCTTAGTCTCGGGCAGTTATACTTAAGCATGAACATTGACGACAAACTGGAAGGATTG  420
                                  M  N  I  D  D  K  L  E  G  L        10

421   TTTCTTAAATGTGGCGGCATAGACGAAATGCAGTCTTCCAGGACAATGGTTGTAATGGGT  480
      F  L  K  C  G  G  I  D  E  M  Q  S  S  R  T  M  V  V  M  G     30

481   GGAGTGTCTGGCCAGTCTACTGTGTCTGGAGAGCTACAGGATTCAGTACTTCAAGATCGA  540
      G  V  S  G  Q  S  T  V  S  G  E  L  Q  D  S  V  L  Q  D  R     50

541   AGTATGCCTCACCAGGAGATCCTTGCTGCAGATGAAGTGTTACAAGAAAGTGAAATGAGA  600
      S  M  P  H  Q  (E) I  L  A  A  (D) (E) V  L  Q  (E) S  (E) M  (R)  70

601   CAACAGGATATGATATCACATGATGAACTCATGGTCCATGAGGAGACAGTGAAAAATGAT  660
      Q  Q  (D) M  I  S  H  (D) (E) L  M  V  H  (E) (E) T  V  (K) N  (D)  90

661   GAAGAGCAGATGGAAACACATGAAAGACTTCCTCAAGGACTACAGTATGCACTTAATGTC  720
      (E) (E) Q  M  (E) T  H  (E) (R) L  P  Q  G  L  Q  Y  A  L  N  V  110

721   CCTATAAGCGTAAAGCAGGAAATTACTTTTACTGATGTATCTGAGCAACTGATGAGAGAC  780
      P  I  S  V  K  Q  E  I  T  F  T  D  V  S  E  Q  L  M  R  D     130

781   AAAAAACAAATCAGAGAGCCAGTAGACTTACAGAAAAAGAAGAAGCGGAAACAACGTTCT  840
      K  K  Q  I  R  E  P  V  D  L  Q  K  K  K  K  R  K  Q  R  S     150

841   CCCGCAAAAATCCTTACAATAAATGAGGATGGATCACTTGGTTTGAAAACCCCTAAATCT  900
      P  A  K  I  L  T  I  N  E  D  G  S  L  G  L  K  T  P  K  S     170

901   CACGTTTGTGAGCACTGCAATGCTGCCTTTAGAACGAACTATCACTTACAGAGACATGTC  960
      H  V  C  E  H  C  N  A  A  F  R  T  N  Y  H  L  Q  R  H  V     190

961   TTCATTCATACAGGTGAAAAACCATTTCAATGTAGTCAATGTGACATGCGTTTCATACAG  1020
      F  I  H  T  G  E  K  P  F  Q  C  S  Q  C  D  M  R  F  I  Q     210

1021  AAGTACCTGCTTCAGAGACATGAGAAGATTCATACTGGTGAAAAACCATTTCGCTGTGAT  1080
      K  Y  L  L  Q  R  H  E  K  I  H  T  G  E  K  P  F  R  C  D     230

1081  GAATGTGGTATGAGATCCATACAAAAATATCATATGGAAAGGCATAAGAGAACTCATAGT  1140
      E  C  G  M  R  S  I  Q  K  Y  H  M  E  R  H  K  R  T  H  S     250

1141  GGAGAAAAACCTTACCAGTGTGAATACTGTTTACAGTATTTTTCCAGAACAGATCGTGTA  1200
      G  E  K  P  Y  Q  C  E  Y  C  L  Q  Y  F  S  R  T  D  R  V     270

1201  TTGAAACATAAACGTATTGGCCATGAAAATCATGACAAAAAACTAAATACATGTGCCATG  1260
      L  K  H  K  R  I  G  H  E  N  H  D  K  K  L  N  T  C  A  H     290
```

```
1261  AAAGGTGGCCTTCTGCGCTCTGAGGAAGATTCTGGCTTTTCTACATCACCAAAAGACAAC  1320
      K  G  G  L  L  R  S  E  E  D  S  G  F  S  T  S  P  K  D  N    310

1321  TCACTGCCAAAAAAGAAAAGGCAGAAAACGGAGAAAAAATCATCTGGAATGGACAAAGAG  1380
      S  L  P  K  K  K  R  Q  K  T  E  K  K  S  S  G  M  D  K  E    330

1381  AGTGCTTTGGACAAATCTGACCTGAAAAAAGACAAAAATGATTACTTGCCTCTTTATTCT  1440
      S  A  L  D  K  S  D  L  K  K  D  K  N  D  Y  L  P  L  Y  S    350

1441  TCAAGTACTAAAGTAAAAGATGAGTATATGGTTGCAGAATATGCTGTTGAAATGCCACAT  1500
      S  S  T  K  V  K  D  E  Y  M  V  A  E  Y  A  V  E  M  P  H    370

1501  TCGTCAGTTGGGGGCTCGCATTTAGAAGATGCGTCAGGAGAAATACACCCACCTAAGTTA  1560
      S  S  V  G  G  S  H  L  E  D  A  S  G  E  I  H  P  P  K  L    390

1561  GTTCTCAAAAAAATTAATAGTAAGAGAAGTCTGAAACAGCCACTGGAGCAAAATCACACA  1620
      V  L  K  K  I  N  S  K  R  S  L  K  Q  P  L  E  Q  N  H  T    410

1621  ATTTCACCTTTATCCACATATGAAGAGCGAAAGTTTCAAAGTATGCTTTTGAACTTGTGG  1680
      I  S  P  L  S  T  Y  E  E  R  K  F  Q  S  M  L  L  N  L  W    430

1681  ATAAACAGGCTTTACTGGACTCAGAAGGCAATGCTGACATTGATCAGGTTGATAATTTGC  1740
      I  N  R  L  Y  W  T  Q  K  A  M  L  T  L  I  R  L  I  I  C    450

1741  AGGAGGGCCCAGTAAACCTGTGCATAGTAGTACTAATTATGATGATGCCATGCAGTTTTT  1800
      R  R  A  Q  *

1801  GAAGAAGAAGCGGTATCTTCAAAGCAAGTAACAACAGCAGGGAATATGCGCTGAATGTGG  1860

1861  GTACCATACGTTCTCAGCCTTCTGTAACACAAGCAGCTGTGGCAAGTGTCATTGATGAAA  1920

1921  GATCCACGGCATCCATATTAGAGTCACAGGCACTGAATGTGGAGATTAAGAGTAATCATG  1980

1981  ACAAAAATGTGTTATTCCAGATGAGGTACTGCAGACTCTGTTGGATCATTATTCCCACAA  2040

2041  AGCTAATGGACAGCATGAGATATCCTTCAGTGTTGCAGATACTGAAGTGACTCTAGCATA  2100

2101  TCAATAAATTCTTCAGAAGTCCAGAGTCACCCCGTCAGAGAATGTTGATCAAGCTCCCAA  2160

2161  GCATCCTCATCAGATAAAGCCAACATGTTGCAGGAATACTCCAAGTTTCTGCAGCAGGCT  2220

2221  TTGGACAGAACTAGCCAAAATGATGCCTATTTGAATAGCCCGAGCCTTAACTTTGTGACT  2280

2281  GATAACCAGACCCTCCCAAATCAGCCAGCATTCTCTTCCATAGACAAGCAGGTCTATGCC  2340

2341  ACCATGCCCATCAATAGCTTTCGATCAGGAATGAATTC  2378
```

*Figure 3* DNA sequence and predicted amino acid sequence of htβ. The amino acid sequences predicted from the longest reading frame of htβ are shown. The zinc finger motifs are shown with an underline; the highly acidic segment that occurs amino terminal to the zinc fingers is denoted by a double underline, and the charged amino acids are circled. Both basic segments that are located before and after the zinc fingers are shown with broken underline. The six in-frame stop codons begins at the "*".

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| consensus | T | G | E | K | P | Y | X | C | $X_{2-4}$ | C | $X_3$ | - | F | -- | $X_5$ | -- | L | $X_2$ | H | $X_{3-4}$ | - | H | | | | | | | |
| htβ 173-193 | | | | | | | | C | E | H | C | N | A | A | F | R | T | N | Y | H | L | Q | R | H | V | F | I | | H |
| htβ 194-221 | T | G | E | K | P | F | Q | C | S | Q | C | D | M | R | F | I | Q | K | Y | L | L | Q | R | H | E | K | I | | H |
| htβ 222-249 | T | G | E | K | P | F | R | C | D | E | C | G | M | R | S | I | Q | K | Y | H | M | E | R | H | K | R | T | | H |
| htβ 250-278 | S | G | E | K | P | Y | G | C | E | Y | C | L | Q | Y | F | S | R | T | D | R | V | L | K | H | K | R | I | G | H |

*Figure 4* Amino acid sequence of four zinc finger motifs of htβ protein aligned to a zinc finger motif consensus. The top line shows the zinc finger structural motif consensus sequence together with the interfinger linker consensus sequence (34). The four htβ zinc finger motifs are shown below the consensus. The amino acids that are highly conserved are boxed. "X" with subcript indicates the number of amino acid residues between the conserved ones.

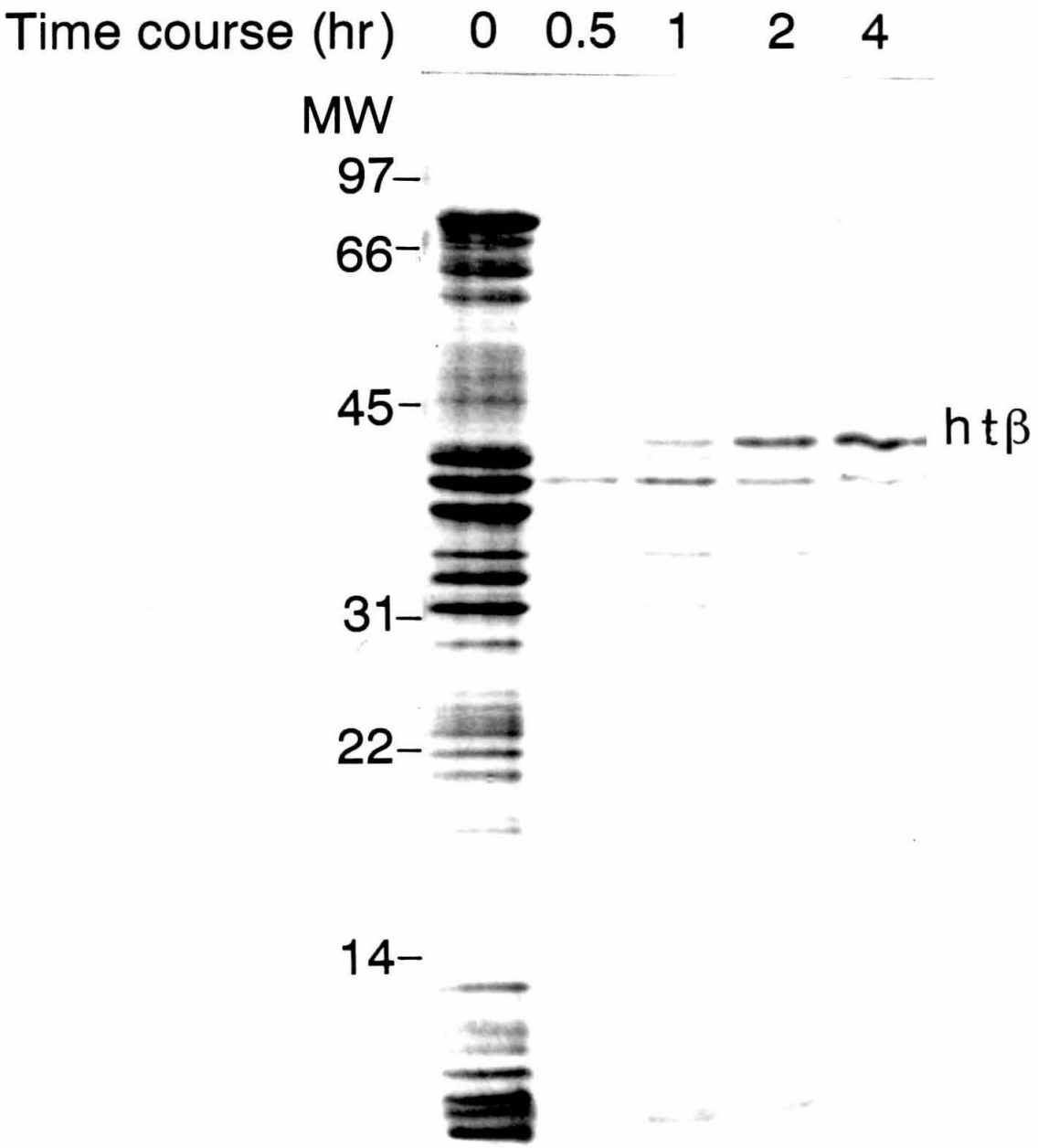
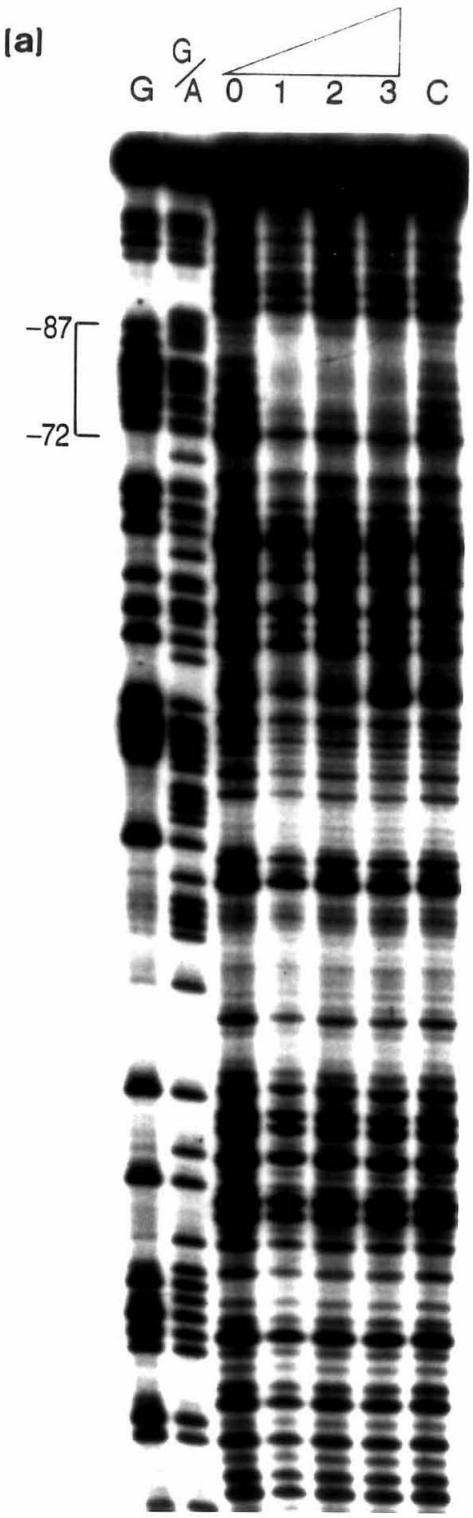

Time course (hr)
0
0.5
1
2
4
MW
97–
66–
45–
31–
22–
14–
htβ

*Figure 5* Accumulation of htβ protein. Electrophoresis patterns of protein (visualized by coomassie brilliant blue staining) are shown for samples collected immediately (0), and 0.5, 1, 2 and 4 hr after IPTG induction of expression of target DNA. Samples of culture (100 μl) were used for electrophoresis through a 12% polyacrylamide gel in the presence of sodium dodecyl sulfate.


(a)
G
G/A
0
1
2
3
C
-87
-72

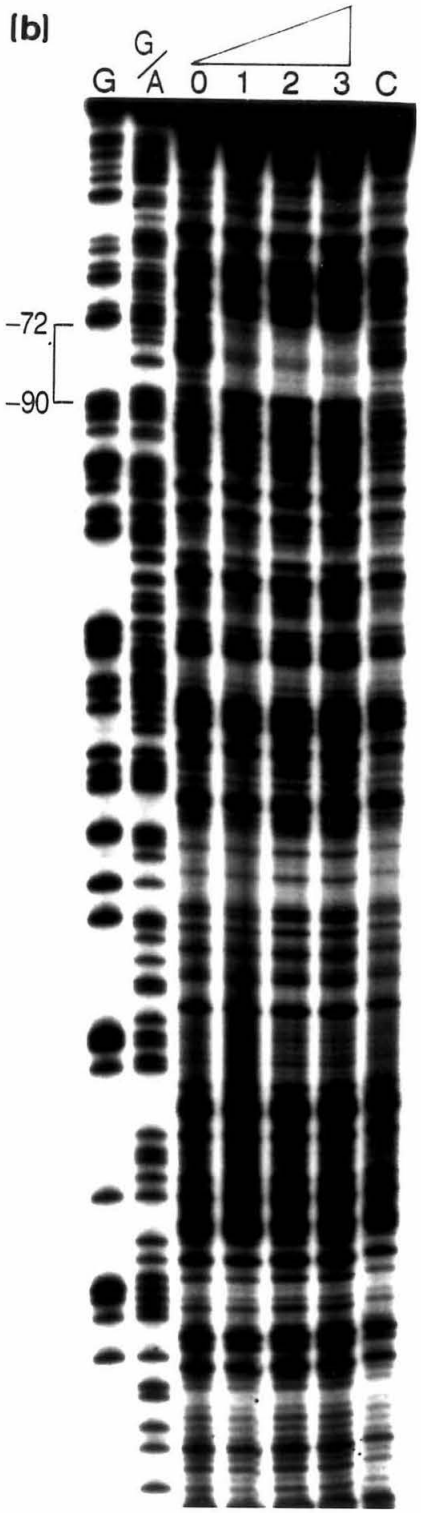

(b)
G
G/A
0 1 2 3
C
-72
-90

*Figure 6* DNase I footprinting assay of coding (a) and noncoding (b) strands of human TCR Vβ8.1 promoter. G and G/A, Maxam and Gilbert sequencing ladder. (a) +55 to -121 bp region of the Vβ8.1 promoter, showing a binding site at -72 to -87 bp. (b) -384 to +25 bp region of the Vβ8.1 promoter, showing a binding site at -72 to -90 bp. In both (a) and (b), lane 0, no protein extract. The other lanes contained extract from bacteria containing the pET3a-htβ expression plasmids (lanes 1-3) or the control pET3a vector alone (lane c) as follows: lane 1, 0.5 μg; lane 2, 1 μg; lane 3, 4 μg; lane c, 4 μg.

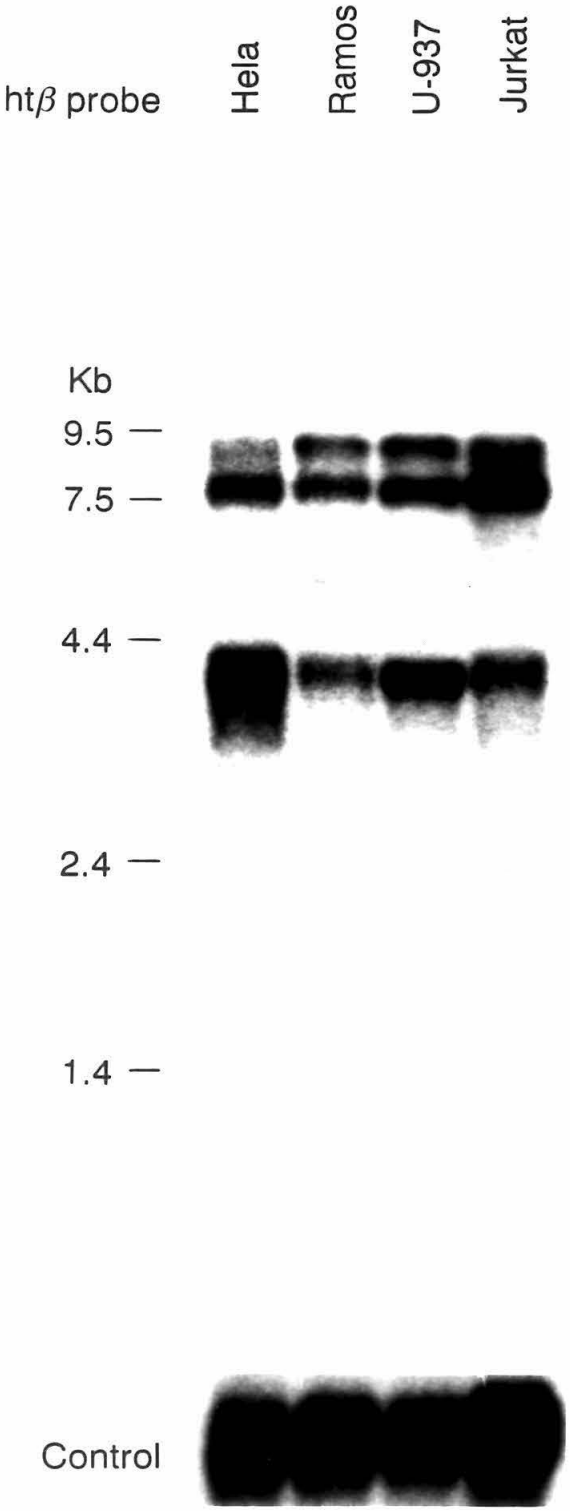

htβ probe
Hela
Ramos
U-937
Jurkat
Kb
9.5
7.5
4.4
2.4
1.4
Control

*Figure 7* Northern blot analysis of poly $(A)^+$ RNAs from various cell lines. Poly $(A)^+$ RNAs (10 μg) were loaded on each lane. The blot was hybridized with labeled 1.8 kb insert of htβ16.

## Chapter 2

Transcriptional Regulation of T Cell Receptor Genes Mediated by a Novel CACCC Box Binding Protein Htβ

**ABSTRACT**

Htβ is a zinc finger containing protein that binds to the human T cell receptor (TCR) Vβ8.1 gene promoter. Htβ also can bind to the mouse T cell receptor gene α silencer. The comparison of htβ binding sites between the human TCR Vβ8.1 promoter and the mouse TCR gene α silencer reveals a core sequence of the CACCC box. Gel-shift assay of five repeats of the CACCC box with bacterially expressed htβ protein indicates that htβ can bind to the CACCC box. Gel-shift assay analysis of the CACCC box with nuclear extracts from various cell lines reveals four common bands in T cell, B cell, monocyte and Hela cell lines; one extra band is seen in Hela cell extracts. CAT assay analysis indicated the CACCC box is essential for efficient transcription of the Vβ 8.1 promoter. Cotransfection with a htβ expression plasmid and a reporter vector show that htβ can activate the human TCR Vβ8.1 gene transcription. Htβ also is able to counteract the silencing effect of the murine TCR α silencer. Htβ may have an interaction with the cAMP response element binding protein (CREB) to negatively regulate human TCR Vβ8.1 gene transcription in Hela cells, and that negative effect was not significant in Jurkat T cells. The CACCC box has been found in almost all Vβ8.1 gene subfamily members, and in both TCR α and β enhancers in human and mouse. These results suggest that the CACCC box binding protein may have an important regulatory function on the immune system.

## INTRODUCTION

Transcriptional regulation of gene expression in mammals is mediated in part by sequence-specific factors that bind to cis-elements such as promoters, enhancers and silencers (1, 2). Many of these factors have been characterized and categorized into four major gene families according to their DNA binding domain structures: the homeodomain, zinc finger, leucine zipper and helix-loop-helix. Some of the factors have been shown to be directly involved in regulation of mRNA synthesis. Promoters and enhancers usually have sequence binding sites for various transcription factors. These factors may act together to regulate specific transcription for each gene in mammalian cells. So it is very important to know how these factors interact with DNA sequence and how these DNA binding proteins interact with each other. Although a lot of transcription factors have been characterized and well studied, the transcription factor which binds to the CACCC box has not been isolated. The CACCC box was first described in the β-globin gene promoter region and was required for efficient and accurate β-globin gene expression. Mutation at the CACCC box region strongly reduced transcription, suggesting that the CACCC box is an important promoter element (3). The CACCC box was not only observed in most of the globin gene family, but also in the erythroid transcription factor GATA-1 promoter (4), slow/cardiac troponin C gene enhancer (5), rat gastric $H^+/K^+$-ATPase β gene promoter (6), glucocorticoid receptor gene promoter (7), and mouse immunoglobulin germ-line $C\gamma_1$ gene promoter (8). Very likely the CACCC box binding protein functions as a general and important transcription factor for gene expression. In addition to activating transcription through interacting with DNA sequences, the CACCC box binding proteins interact with other transcription factors to carry out more functions. A human T cell receptor Vβ8.1 gene promoter binding protein, htβ, was characterized in Chapter 1. In this chapter, we report the

further characterization of htβ as a possibly a long-sought CACCC binding protein and present functional analysis of htβ protein on T cell receptor gene expression.

## MATERIAL AND METHODS

### *Oligonucleotides*

The primers for amplifying the murine TCR gene α silencer I (9) from mouse genomic clone TA4.1 (10) are:

Primer 1 5' CTCGAGCAACATGTGGGAAGA 3'

Primer 2 5' CAGCTGTCTGTGTGATTCTTG 3'

The double-stranded probe used in gel-shift analysis was produced by the following two oligonucleotides:

```
5' GATCTGGGGGTGGGGTGGGGGTGGGGTGGGGGTGGGG 3'
    3' ACCCCCACCCCACCCCCACCCCACCCCCACCCCCTAG 5'
```

The CACCC boxes are indicated.

The oligonucleotides for gel-shift competition assays have two 17 mer GAL4 factor binding sites (11). The sequence is:

```
5' CGGAGTACTGTCCTCCGCGGAGTACTGTCCTCCGCTGC 3'
3' GCCTCATGACAGGAGGCGCCTCATGACAGGAGGCGACG 5'
```

### *Plasmid construction*

Expression plasmids. The htβ full length cDNA (Chapter 1) was subcloned into the EcoRI site of pSG5, a eukaryotic expression vector (Stratagene) in which the htβ full length cDNA is under the control of the SV40 early promoter.

Reporter plasmids. The human Vβ8.1 promoter region (12) of various lengths was subcloned into the pCAT enhancer plasmid (Promega) that has a CAT gene and a SV40 enhancer, but lacks the promoter element. For pVβ8.1CAT, the MseI-BstNI (-91

to +17) fragment of human Vβ8.1 was filled-in at the ends using Klenow DNA polymerase and then ligated to XbaI digested and filled-in pCAT enhancer plasmid. pΔ Vβ8.1CAT was constructed by filling-in a HaeIII fragment (-71 to +14) and subcloning it into the pCAT enhancer plasmid as pVβ8.1CAT. The oligonucleotide containing five repeats of the CACCC was ligated to HindIII digested and filled in pΔVβ8.1CAT, placing the five repeats of the CACCC upstream of the Vβ8.1 promoter HaeIII fragment. This new plasmid is named pOΔVβ8.1CAT.

The mouse TCR α gene silencer I fragment (9) was made by PCR amplification of mouse genomic clone TA 4.1 (10) and confirmed by DNA sequence analysis. The 309 bp mouse TCR silencer I fragment was ligated to either BglII or XbaI cleaved and filled-in pCAT Control plasmid (Promega) that has the SV40 promoter, enhancer and a CAT gene. The mouse TCR α silencer was placed either upstream of the SV40 early promoter in the case of pControl digested with BglII (named pControl sil I 5') or downstream of the SV40 enhancer when digested with XbaI (named pControl sil I). All the plasmids constructs were confirmed by DNA sequence analysis.

*Bacterial expression of htβ gene, gel-shift assay and DNase I footprinting assay*

Bacterial expression of htβ gene, gel-shift assay and DNase I footprinting assay were performed as described in Chapter1.

*Nuclear extract preparation and gel-shift assay*

Hela cells were grown in DMEM media containing 10% FCS (fetal bovine serum). U-937, Ramos and Jurkat cells were grown in RPMI 1640 media containing 10% FCS. Cells were harvested in log phase and nuclear extracts were prepared according to Schreiber et al (13). Gel-shift assays were done using the conditions described by Fan and Maniatis (14). Nuclear extracts (10 μg) were incubated with 2.5 ng of $^{32}$P-labeled probe and 2.5 μg of poly (I-C) (Boehringer Mannheim) on ice for 20

minutes. The reaction mixtures were loaded in a 6% polyacrylamide/0.5xTBE (Tris-borate-EDTA buffer) gel.

*Transfection and cAMP induction*

Hela cells were transfected with plasmid DNA using liposomes (15). Briefly, exponentially growing cells were plated in 6-well tissue culture dishes at $5x10^5$ cells/well and grown overnight. Plasmid DNA was mixed with 20 μg of Lipofectin reagent (GIBCO BRL) in 1 ml serum-free DMEM. 1 ml DNA/liposome complex was added directly to the cells. Incubation for 3 hrs at 37°C was followed with the addition of 1 ml DMEM containing 20% FCS. Cells were harvested 48 hrs after transfection for CAT assay analysis.

The DEAE-dextran method (16) was used for transfection of the Jurkat cells. Cells ($5x10^6$) were incubated in 1 ml of TBS (Tris-buffered saline) containing plasmid DNA and 500 μg/ml DEAE-dextran (Promega) for 30 minutes at room temperature. The cells were then washed with TBS and incubated in 20 ml RPMI 1640 media containing 10% FCS for 48 hrs. As a control for differences in transfection efficiencies, 0.5-1 μg of a β-galactosidase gene containing plasmid (CMV β-gal, a kind gift from Ebrahim Zandi, Department of Chemistry and Chemical Engineering, Caltech) was included in each transfection. For cAMP induction, forskolin (Sigma, final 10 μM) and 3-butyl-1-methyl-xanthine (IMX) (Sigma, final 100 μM) were added to the culture media right after transfection (17).

*Chloramphenicol Acetyl-transferase (CAT) Assay and β-Galactosidase Assay*

Cells were harvested by scraping with a rubber policeman (Hela cells), centrifuged, and then washed once in PBS (phosphate buffered saline solution). The cells are resuspended in 100 μl 0.25 M Tris-HCl, pH7.4, and subjected to three cycles of freeze-thaw lysis in liquid nitrogen. The subsequent assays were done according to

Gorman et al. (18) using 0.2 μCi $^{14}$C-chloramphenicol (Amersham) and 25 μg acetyl-coenzyme A (Sigma) for each reaction. The acetylated and nonacetylated forms of chloramphenicol were excised from the thin-layer chromatography, and the amount of radioactivity was determined by liquid scintillation counting. β-galactosidase assays were done as described (19) using O-nitro-phenyl-β-D-galactopyranoside (Sigma) as a substrate. The amount of cell extracts used for CAT assays were optimized according to β-galactosidase assay results.

## RESULTS

### *Htβ binds to the CACCC motif*

In Chapter 1, the binding site for htβ protein in the human TCR Vβ8.1 promoter has been characterized. On the basis of DNase I footprinting with E. coli expressed htβ protein, the sequence of the site is 5' GAAGTTGGGGGTGGTG 3' (fig. 1). Other T cell receptor gene regulatory regions have been searched for the presence of a htβ binding site. The mouse TCR gene α silencer I has a region of sequence (9) that is highly similar to the htβ binding site in the Vβ8.1 gene promoter. Crude lysates of E. coli strain BL21 (DE3) containing pET3a-htβ were tested for DNA binding activity by DNase I footprinting with the mouse TCR α silencer I fragment. The extracts prepared from BL21 (DE3) (pET3a-htβ) can protect the 163-193 bp region of the mouse TCR α silencer from digestion by DNase I (fig. 2).

When the htβ binding sites in both human Vβ8.1 and mouse TCR gene α silencer I were compared (fig. 3), a consensus sequence TGGGGGTGG was found, which has the complement of the well-known CACCC box (3) in the center. To find out whether only this CACCC motif is necessary for htβ protein binding, double-stranded oligonucleotides containing five repeats of the CACCC box were synthesized, annealed,

and labeled with $^{32}P$ as probes for gel-shift analysis. The probe can form a DNA-protein complex with extracts prepared from BL21 (DE3) (pET3a-htβ), but not BL21 (DE3) (pET3a) (fig. 4). This result clearly shows that CACCC is the core sequence for htβ binding. Very likely htβ is one of the long-sought CACCC box binding proteins.

In addition, nuclear extracts from Hela cells, U-937 monocytes, Ramos B cells and Jurkat T cells were tested with the same probe containing five CACCC boxes in a gel-shift assay (fig. 5). Four common bands (band 1-4) are visible with all nuclear extracts tested. However, an additional fast-moving band (band 5) is only seen with extracts from Hela cells. To demonstrate the specificity of binding to the CACCC box, an excessive amount of unlabeled CACCC containing oligonucleotide or GAL4 binding site containing oligonucleotides which have no homology to the CACCC box were added to the reaction mixture of $^{32}P$-labeled CACCC containing oligonucleotides and either Hela cell or Jurkat cell nuclear extracts (fig. 6). All bands in the Hela cell extract assay are efficiently competed at 20-50 ng of cold CACCC containing fragment, but not with 50 ng of GAL4 binding site containing oligonucleotide. In Jurkat T cells, Band 1, 2 and 4 are competed with specific CACCC box containing fragment, while band 3 is not competed at 50 ng unlabeled CACCC containing fragment, which is a 20-fold molar excess of competitor.

*CACCC motif is important for human TCR Vβ8.1 gene expression*

To study the importance of the CACCC box in TCR Vβ8.1 transcription, various Vβ8.1 promoter constructs were made and analyzed for expression in both Hela and Jurkat T cells (fig. 7). pVβ8.1CAT, which contains the natural CACCC box in the Vβ 8.1 gene promoter, has significant CAT activity. But in the case of pΔVβ8.1, which has a deletion of the upstream CACCC, the CAT activity is dramatically decreased to about 5% and 2% of the transcriptional activity of pVβ8.1 CAT in Hela and Jurkat cells respectively. The addition of five repeats of the CACCC box upstream of the Vβ8.1

promoter of pΔ8.1 CAT plasmid increased CAT activities greatly, 59% and 100% of the transcriptional activity of pVβ8.1CAT was restored in Hela and Jurkat cells respectively.

*Htβ activates human TCR Vβ8.1 gene expression and antagonizes the function of the mouse TCR α silencer*

To study the regulatory function of htβ protein, the full length cDNA of htβ was subcloned into an eukaryotic gene expression vector under the control of the SV40 early promoter (pSG5). The expression plasmids were cotransfected into Hela cells with the pVβ8.1CAT plasmid that has a Vβ8.1 promoter containing the CACCC box followed by a CAT gene. Due to the presence of endogenous htβ protein, cotransfection with the ht β expression plasmid and reporter plasmid pVβ8.1CAT did not result in significantly increased CAT activity, compared to transfection with reporter plasmid alone. However, the 2-fold increase in CAT activity is observed consistently (fig. 8). Htβ protein is able to activate transcription of human Vβ8.1 gene.

Besides binding to the Vβ8.1 promoter, htβ also can bind to another TCR regulatory sequence, the mouse TCR α gene silencer I. The regulation of the mouse TCR α silencer I by htβ protein was investigated. The presence of the mouse α silencer I downstream of the SV40 enhancer (pControl sil) or upstream of the SV40 promoter (pControl sil 5') results in decreased transcription by approximately 40% and 67% respectively (Fig. 9), while cotransfection of pControl sil or pControl sil 5' with pSG5-ht β restored the transcription activity to 97% or 61%. This indicates that htβ can reverse the silencing function of the mouse α silencer.

*Effects of cAMP on Vβ8.1 gene expression*

There is a cAMP response element (CRE) site in the human TCR Vβ8.1 gene (fig. 1). This DNA sequence has been shown to mediate cAMP responsiveness of rat somatostatin (20). Studies were carried out to investigate the effect of cAMP on Vβ8.1

gene expression. Both Hela cells and Jurkat T cells were treated with forskolin and 3-butyl-1-methyl-xanthine (IMX) for cAMP induction right after transfection. The transcription activity of Vβ8.1 was decreased to 20% in Hela cells when induced with cAMP, while transcription of Vβ8.1 in Jurkat cells was reduced to 61% upon cAMP induction (fig. 10). cAMP induction has little effect on pControl plasmid which carries the SV40 promoter and enhancer. These results suggest that the Vβ8.1 gene was down-regulated through cAMP. However the down-regulation was more pronounced in Hela cells than in Jurkat T cells.

## DISCUSSION

The htβ gene was first isolated with the belief that it was a T cell specific transcription factor. Expression analysis of the htβ gene in cell lines from different lineages makes this unlikely (Chapter 1). In addition to binding to the human TCR Vβ 8.1 promoter, htβ also binds to another T cell regulatory sequence, the mouse T cell receptor α gene silencer (fig. 2). The comparison of the binding sites in both Vβ8.1 and the TCR α silencer I reveals that the consensus sequence for htβ binding is nine nucleotides. This consensus sequence is centered with a reverse oriented CACCC box, which was first characterized as an important promoter element for globin gene expression (3). Oligonucleotides containing five repeats of the CACCC box can specifically form a DNA-protein complex with bacterially-expressed htβ protein in a gel-shift assay. Gel-shift assay analysis with the same CACCC containing probe and nuclear extracts from different cell lines reveal a number of bands. Four common bands are observed in nuclear extracts from Hela, U-937 monocyte cell line, Ramos B cell line and the Jurkat T cell line, which may partially explain the multiple mRNA species seen in Northern blot analysis of these four cell lines with the htβ probe (Chapter 1). No

specific band is observed in nuclear extracts from T cells. Interestingly, an extra fast-moving band is detected in nuclear extracts from Hela cells, and not from the three hematopoietic cells. Whether this CACCC binding protein can regulate epithelial-specific gene expression is not known. Since the htβ gene is single copy (Chapter 1), different htβ protein forms may result from alternative RNA splicing. The question of possible different forms of htβ that can regulate gene expression is unanswered at this time. Also, gel-shift assay comparison of the CACCC containing probe with either in vitro translated protein from full length htβ cDNA or cell nuclear extracts may tell us which band in the gel-shift assay with nuclear extracts corresponds to htβ protein encoded by the gene that we have cloned.

To find out how many genes are possibly regulated by the htβ protein, both GGGTG and CACCC boxes were searched among different genes. Examination of sequence information from 14 mouse Vβ and 5 human Vβ gene promoters (21, 22) reveals that most of the Vβ8 family members of both human and mouse have the GGGTG box. Only the mouse Vβ8.1 gene and the human Vβ8.4 pseudogene have modified GGGTG boxes (fig. 11). Among the other Vβ members examined, only mouse Vβ7 has a GGGTG box. In addition to Vβ genes, other TCR regulatory sequences also have GGGTG or CACCC boxes, including TCR α and β enhancers of both human and mouse (23-26), and mouse TCR gene α silencer I and II. TCR silencer I has four continuous repeats of the GGGTG box and one CACCC box (fig. 12). Functional analysis will be necessary to determine whether these GGGTG or CACCC boxes bind to transcription factors and activate gene expression. The CACCC box is not found in the TCR γ enhancer (27). Although there is a CACCC box in the TCR δ enhancer, DNase I footprinting assay shows no binding activity to that site (28). We believe that htβ protein may act as a major activator for TCR gene expression by binding to the GGGTG boxes in the TCR α silencer to reverse the silencer activity in αβ T cells.

Comparison of htβ gene expression in αβ versus γδ T cells may provide important information about regulation of TCR genes during ontogeny.

In addition to T cell receptor genes, the CACCC box was first described in the globin gene promoter, and later was found in regulatory sequences of many genes (fig. 13). Most of these CCCAC boxes are functional for the transcriptional activation of these genes. The general impression from fig. 11-fig. 13 is that the GGGTG box is more frequently used in TCR genes, and globin and other genes often have the reverse orientation of GGGTG (CACCC) for their transcriptional activation. The orientation of the CACCC box may not have much influence on transcription. In the rat tryptophan oxygenase gene promoter, it has been shown that both the orientation and number of the CACCC box do not affect the gene transcription significantly (29). Deletion of the htβ binding site which has a CACCC box in it dramatically reduces transcription activity, while addition of five repeats of CACCC box to the deleted Vβ8.1 promoter resulted in major increase of the transcriptional activity. The results clearly show that the CACCC box is very important for efficient transcription of the Vβ8.1 promoter. Although there is a CACCC box in the SV40 enhancer, it is not important for transcription in this system. pΔVβ8.1CAT plasmid which has deleted the CACCC box in the Vβ8.1 promoter but carries the CACCC box in the SV40 enhancer only expresses basal transcriptional activity (fig. 7). One study has shown the upstream CACCC box in the human β-globin gene promoter is unable to bind CACCC factor, in contrast to the strong effect of the proximal CACCC box on globin transcription. In spite of the existence of the upstream CACCC box, mutation at the proximal CACCC box greatly decreased globin gene transcription (30). DNase I footprinting assay analysis (fig. 2) indicates htβ binds to the four continuous GGGTG boxes in the mouse α silencer I very well, but significant binding is not observed in the CACCC box located 101 nucleotides away. The sequences flanking the CACCC box may have a significant effect on the ability of its binding to transcription factors.

The importance of the CACCC box in gene transcription was first shown in the study of the β-globin promoter. The C→T transition at the first position of the CACCC box in the β-globin promoter strongly decreased transcription (3). A CACCC box mutation in the slow/cardiac troponin C gene enhancer also resulted in approximately 90% reduction of transcription relative to that observed with the wild-type enhancer (5). Similar results were obtained from the mouse porphobilinogen deaminase gene (31), the human 7S K RNA gene (31) and GATA-1 gene promoters (4). The function of the CACCC box in other T cell receptor genes has yet to be determined.

The CAT assay of cotransfection with an expression plasmid that expresses the ht β protein and a reporter plasmid that has a human Vβ8.1 promoter containing CACCC box indicates that htβ protein can activate Vβ8.1 promoter transcription. Endogenous htβ protein results in transcriptional activity in the absence of additional htβ protein. The observed 2-fold increase in activity was reproducible. Htβ protein can reproducibly counteract the silencing activity of mouse TCR α silencer I (fig. 9). Cotransfection of the htβ expression plasmid with pControl plasmid which has both the SV40 promoter and enhancer did not increase transcription. The antagonizing activity of htβ protein on the TCR gene α silencer may have interesting implications for TCR gene expression in αβ and γδ T cells. We speculate that perhaps in γδ T cells, either htβ protein is not expressed, or the htβ protein is functionally different from that in αβ T cells possibly due to post-translational modification, or the CACCC boxes in the TCR α silencer are not accessible to htβ protein since αβ and γδ T cells may have different chromatin structure at the α/δ locus. Any one of the above reasons may result in the silencing of the α locus in γδ T cells. On the contrary, in αβ T cells the actively expressed htβ protein may bind to the CACCC boxes in both silencer I and II, resulting in the relief of silencer activity, leading to the expression of the TCR α gene. The details of how htβ protein is involved in TCR α gene activity of αβ and γδ T cells may help to explain the differentiation process of these two populations of T cells.

The interaction between the CACCC box binding protein and other transcription factors has been reported. CACCC binding protein is synergistic with NF-F1 erythroid specific transcription factor (33), glucocorticoid receptor (29,34), and IL-4 (8) for particular gene expression. Since a cAMP response element (CRE) is just downstream of the CACCC box in the human Vβ8.1 promoter (fig. 1), induction by cAMP could decrease greatly the transcriptional activity of the Vβ8.1 promoter in Hela cells. This suggests cAMP may negatively regulate transcription from the cAMP response element in Vβ8.1. CREB-2 protein has been shown to cause the repression of CRE-dependent transcription (35). It is not clear whether CREB-2 is responsible for the normally observed repression of Vβ8.1 transcription in Hela cells. Jurkat T cells which actively express the Vβ8.1 gene were not significantly subject to negative regulation as Hela cells by cAMP. An intriguing possibility is that protein-protein interactions between htβ and the CRE binding protein result in regulation of T cell receptor genes.

Another interesting aspect about the CACCC box is that mutation at this binding site can cause disease. Mutation at position -87 (C→T or C→G) in the CAC<u>C</u>C box of the globin promoter greatly decreases transcriptional activity of the β-globin gene, and the natural mutation at that position has been shown to cause thalasemia (36-37). Our study also shows that the CACCC box is very important for T cell receptor gene transcription. Mutation at the CACCC box of TCR promotesr may cause inefficient expression of certain Vβ genes, therefore influencing the immune response carried out by a population of T cells which normally use that Vβ gene product together with a particular Vα gene product to recognize the MHC molecule and foreign peptide.

REFERENCES

1. Mitchell, P. and Tjian, R. (1989) Science 245:371-478.

2. Johnson, P.F. and McKnight, S.L. (1989) Annu. Rev. Biochem. 58:799-839.

3. Dierks, P., van Ooyen, A., Cochran, M.D., Dobkin, C., Reiser, J. and Weissman, C. (1983) Cell 32:695-706.

4. Tsai, S.-F., Strauss, E. and Orkin, S.H. (1991) Genes Dev. 5:919-931.

5. Parmacek, M.S., Vora, A.J., Shen, T., Barr, E., Jung, F. and Leiden, J.M. (1992) Mol. Cell Biol. 12:1967-1976.

6. Maeda, M., Oshuma, K.-I., Tamura, S., Kaya, S., Mahood, S., Reuben, M.A., Lasater, L.S., Sachs, G. and Futai, M. (1991) J Bio. Chem. 266:21584-21588.

7. Govindan, M.V., Pothier, F., Leclerc, S., Palaniswami, R. and Xie, M. (1991) J. Steroid Biochem. Mole. Biol. 40:317-323.

8. Xu, M. and Stavnezer, J. (1992) EMBO J. 11:145-155.

9. Winoto, A. and Baltimore, D. (1989) Cell 59:649-655.

10. Winoto, A., Mjdsness, S. and Hood, L. (1985) Nature 316:832-836.

11. Giniger, E., Varnum, S.M. and Ptashne, M. (1985) Cell 40:767-774.

12. Diamond, D.J., Nelson, H.S. and Reinherz, E.L. (1989) J. Exp. Med. 169:1213-1231.

13. Schreiber, E., Matthias, P., Muller, M.M. and Schaffner, W. (1988) EMBO J. 7:4221-4229.

14. Fan, C.-M. and Maniatis, T. (1990) Genes Dev. 4:29-42.

15. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1987) Current Protocols in Molecular Biology. John Wiley & Sons.

16. Fujita, T., Shibuya, H., Ohashi, T., Yamanishi, K. and Taniguchi, T. (1986) Cell 46:401-407

17. Lee, M.-R., Chung, C.-S., Lion, M.-L., Wu, M., Li, W.-F., Hsueh, Y.-P. and Lai, M.-Z. (1992) J. Immunol. 148:1906-1912.

18. Gorman, C., Moffat, L. and Howard, B. (1982) Mol. Cell Biol. 2:104-1051.

19. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1989). Molecular Cloning, A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

20. Montiminy, M.R., Sevarino, K.A., Wagner, J.A., Mandel, G. and Goodman, R.H. (1986) Proc. Natl. Acad. Sci. USA 83:6682-6686.

21. Anderson, S.J., Chou, H.S. and Loh, D.Y. (1988) Proc. Natl. Acad. Sci. USA 85:3551-3554.

22. Siu, G., Strauss, E.C., Lai, E. and Hood, L. (1986) J. Exp. Med. 164:1600-1614.

23. Ho, I.-C., Yang, L.-H., Morle, G. and Leiden, J.M. (1989) Proc. Natl. Acad. Sci. USA 86:6714-6718.

24. Gottschalk, L.R. and Leiden, J.M.(1990) Mol. Cell Biol. 10:5486-5495.

25. Winoto, A. and Baltimore, D. (1989) EMBO J. 8:729-733.

26. Krimpenfort, P., de Jong, R., Uematsu, Y., Dembic, Z., Ryers, S., von Boehmer, H., Steimetz, M. and Berns, A. (1988) EMBO J. 7:745-750.

27. Kappes, D.J., Browne, C.P. and Tonegawa, S. (1991) Proc. Natl. Acad. Sci. USA 88:2204-2208.

28. Redondo, J.M., Hata, S., Brocklehurst, C. and Krangel, M.S. (1990) 247:1225-1229.

29. Schule, R., Muller, M., Otsuka-Murakami, H. and Renkawith, R. (1988) Nature 332:87-90

30. Mantovani, R., Malgarett, N., Nicolis, S., Giglion, B., Comi, P., Cappellini, N., Bertero, M.T., Caligaris-Cappio, F. and Ottolenghi, S. (1988) 16:4299-4313.

31. Porcher, C., Pitiot, G., Plumb, M., Lowe, S., de Verneuil, H. and Grandchamp, B. (1991) J. Bio. Chem. 266:10562-10569.

32. Kleinert, H., Assert, R. and Benecke, B.-J. (1991) J. Bio. Chem. 266:23872-23877.

33. Frampton, J., Walker, M., Plumb, M. and Harrison, P.R. (1990) Mol. Cell Bio. 10:3838-3842

34. Muller, M., Baniahmed, C., Kaltschmidt, C. and Renkawith, R. (1991) Mol. Endo.

5:1498-1503.

35. Karpiniski, J.M., Morle, G.D., Huggenvik, J., Uhker, M.D. and Leiden, J.M. (1992) Proc. Natl. Acad. Sci. USA in press.

36. Kulozik, A.E., Bellan-Koch, A., Bail, S., Kohne, E. and Klenhauer, E.(1991) Blood 77:2054-2058.

37. Giglioni, B., Comi, P., Ronchi, A., Mantovani, R. and Ottolenghi, S. (1988) Biochem. Biophy. Res. Commun. 164:149-155.

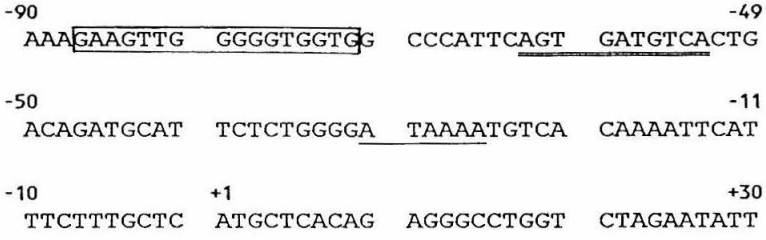

-90
-49
AAAGAAGTTG GGGGTGGTGG CCCATTCAGT GATGTCACTG
-50
-11
ACAGATGCAT TCTCTGGGGA TAAAATGTCA CAAAATTCAT
-10
+1
+30
TTCTTTGCTC ATGCTCACAG AGGGCCTGGT CTAGAATATT

*Figure 1* Nucleotide sequence of the 5' flanking region of the human Vβ8.1 gene between -90 and +30. +1 indicates the putative start site for mRNA synthesis. The htβ binding site is boxed. The putative cAMP response element (CRE) is double-underlined. Also, the TATA box is underlined.

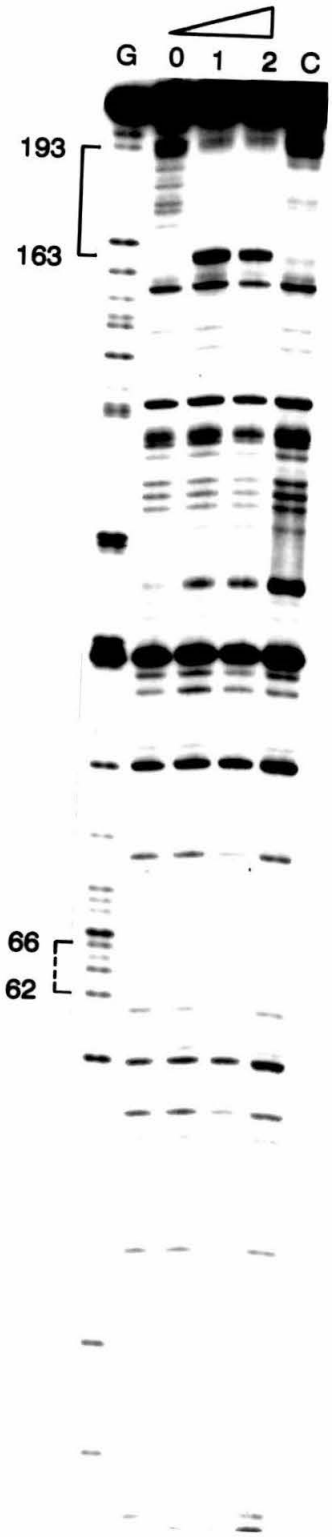

G
0
1
2
C
193
163
66
62

*Figure 2* DNase I footprinting assay of mouse TCR α gene silencer I, 1-309 bp fragment. Lane G, Maxam and Gilbert sequencing ladder. Lane 0, no protein extracts. Lane 1 and 2 contained 1 and 2 μg of protein extracts from bacteria containing the pET3a-htβ expression plasmids. Lane C had 2 μg of the control protein extracts from bacteria containing vector pET3a. The binding site is bracketed (163-193). Also, the CACCC box (62-66) that does not have significant binding activity is bracketed with a broken line.

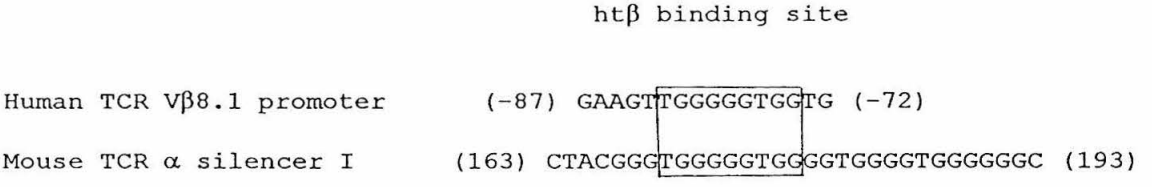

htβ binding site
Human TCR Vβ8.1 promoter (-87) GAAGTTGGGGGTGGTG (-72)
Mouse TCR α silencer I (163) CTACGGGTGGGGGTGGGGTGGGGTGGGGGGC (193)

*Figure 3* Comparison of the htβ binding site in human TCR Vβ8.1 promoter (-87 to -72) and mouse TCR α gene silencer I (163 to 193). The binding site in the Vβ8.1 promoter was defined in Chapter 1, and the binding site in silencer I was determined in *Figure 2*. The homologous regions in both binding sites are boxed.

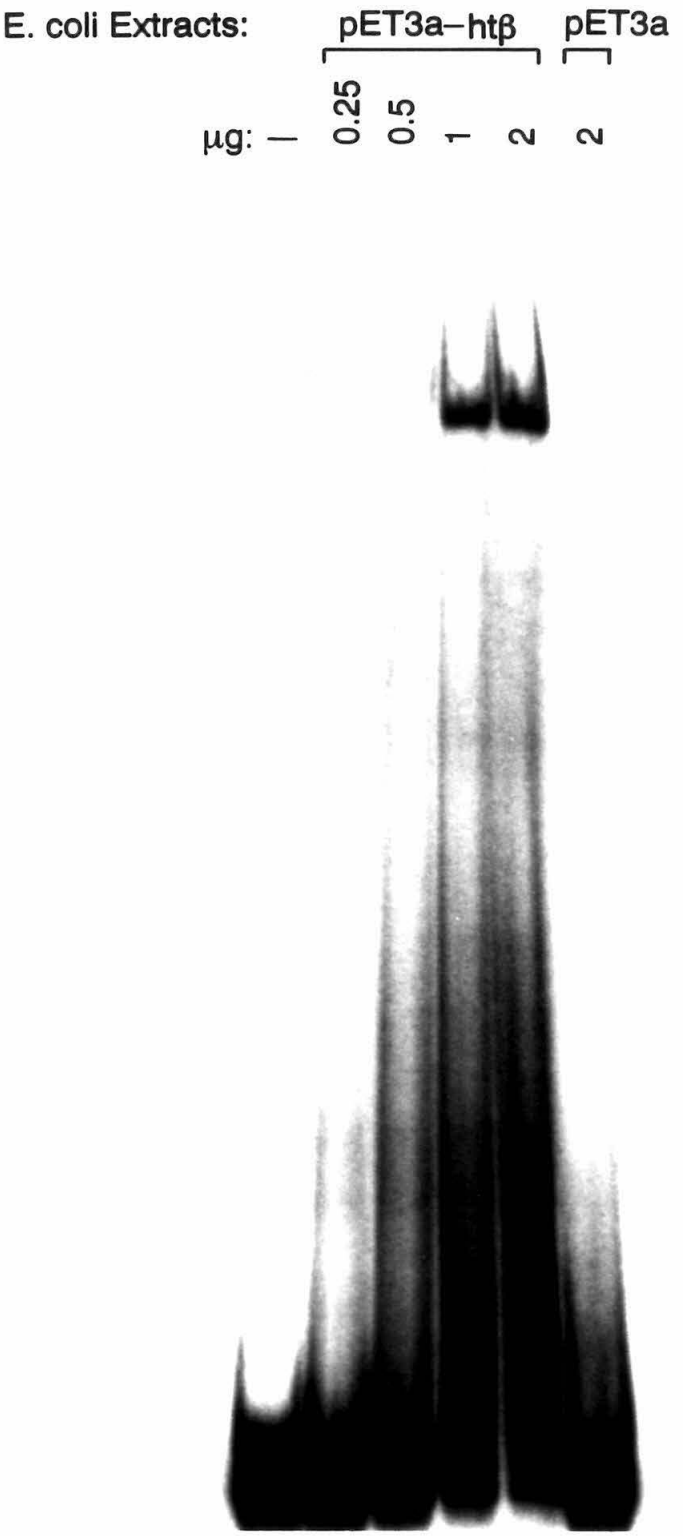

E. coli Extracts:
pET3a–htβ
pET3a
μg: —
0.25
0.5
1
2
2

*Figure 4* Gel-shift assay of synthetic oligonucleotides containing five repeats of the CACCC box with protein extracts from bacteria containing pET3a-htβ plasmids. From left to right lanes, 0, 0.25, 0.5, 1 and 2 μg of protein extracts from bacteria containing pET3a-htβ were used. The rightest lane had 2 μg of protein extracts from bacteria containing pET3a.

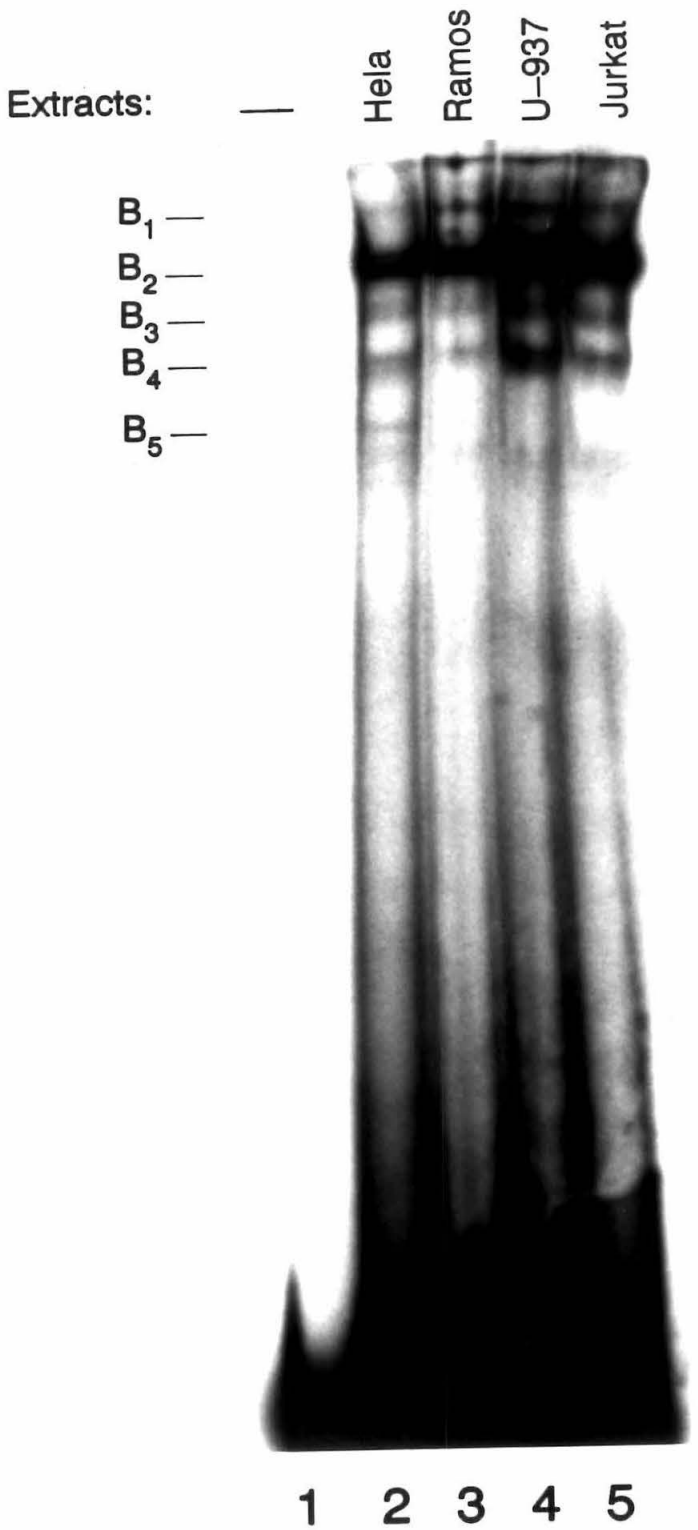

Extracts:
—
Hela
Ramos
U–937
Jurkat
B1
B2
B3
B4
B5
1
2
3
4
5

*Figure 5* Gel-shift assay of synthetic oligonucleotides containing five copies of the CACCC box with nuclear extracts from various cell lines. Lane 1, no protein extracts. Lane 2, 3, 4 and 5 contained 10 μg of nuclear extracts from Hela, Ramos B cell, U-937 monocyte and Jurkat T cell lines respectively. The bands, B1-B5 observed in the gel-shift assay are indicated on the left side.

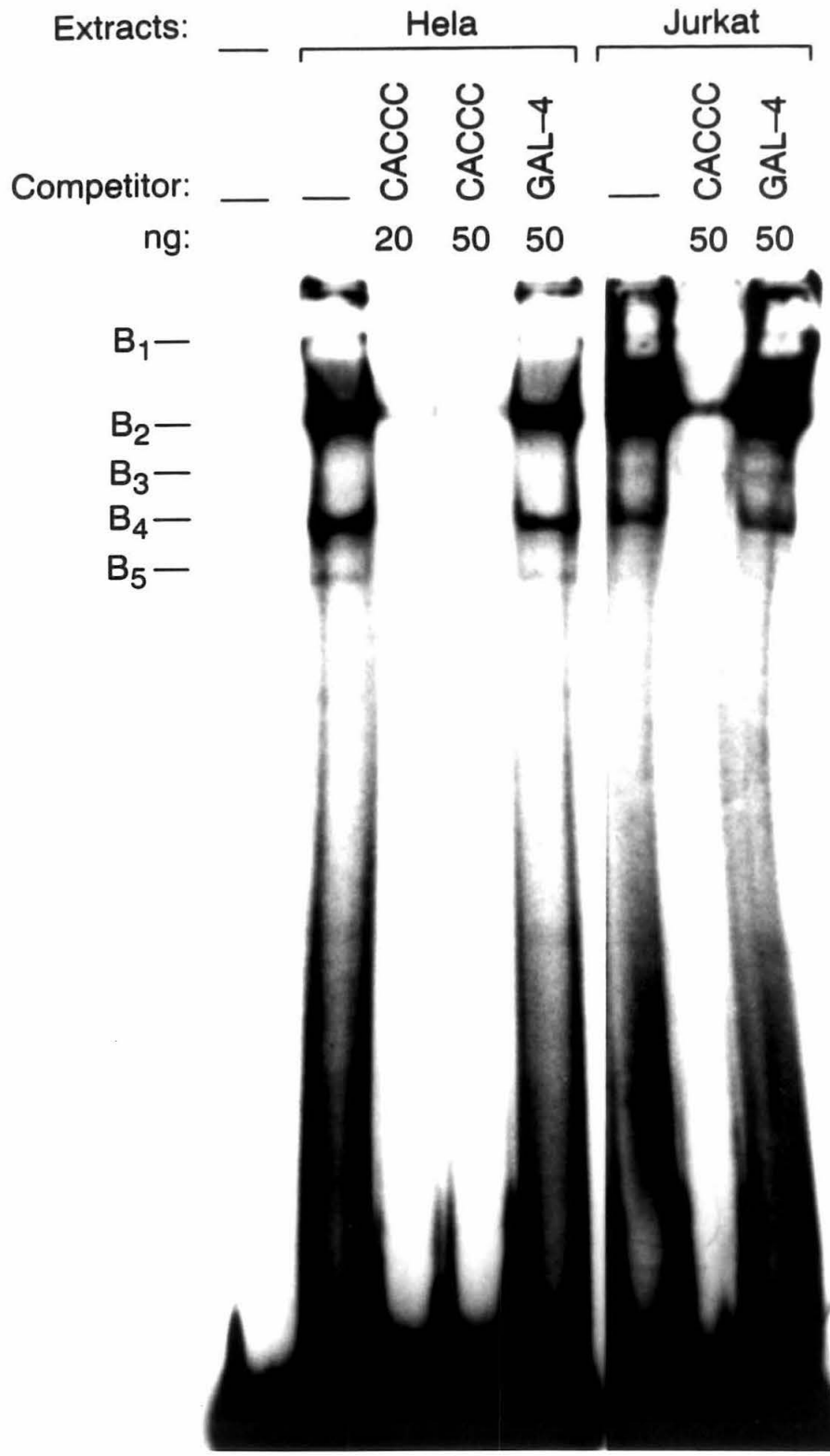

Extracts:
Hela
Jurkat
Competitor:
CACCC
CACCC
GAL-4
CACCC
GAL-4
ng:
20
50
50
50
50
B1
B2
B3
B4
B5

*Figure 6* Identification of nuclear proteins which bind to the CACCC box. $^{32}P$ labeled oligonucleotides containing five repeats of the CACCC box were incubated with 10 μg of nuclear extracts prepared from Hela cells or Jurkat cells. Binding reactions included 20 or 50 ng of the indicated unlabeled competitor oligonucleotides. CACCC, unlabeled oligonucleotides containing five repeats of the CACCC box. GAL-4, unlabeled oligonucleotides containing two copies of the GAL-4 protein binding site.

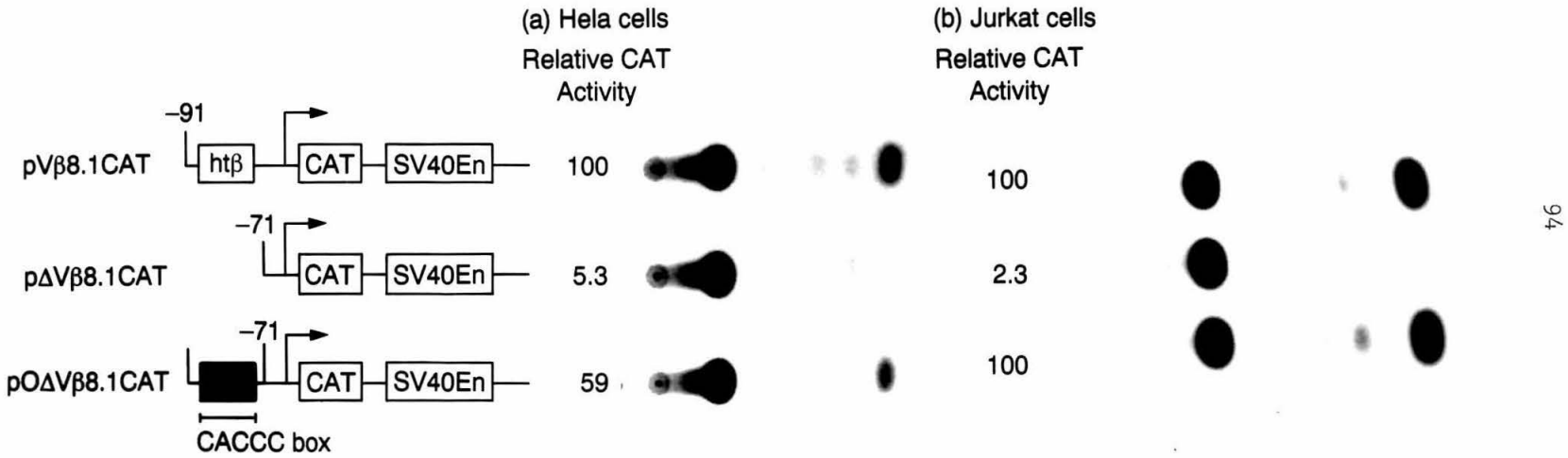
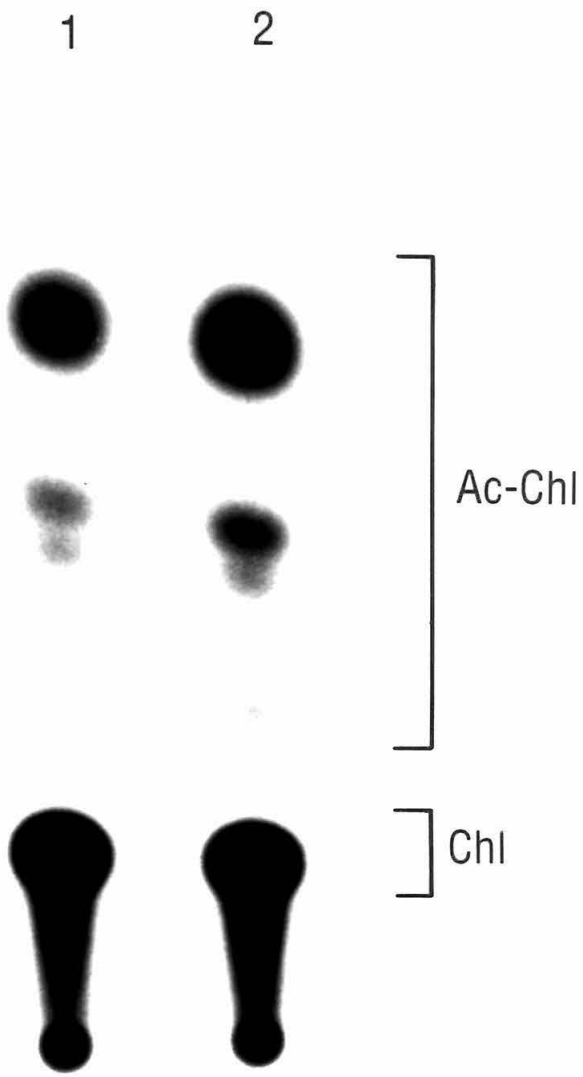

(a) Hela cells
Relative CAT Activity
(b) Jurkat cells
Relative CAT Activity
−91
pVβ8.1CAT
htβ
CAT
SV40En
100
100
−71
pΔVβ8.1CAT
CAT
SV40En
5.3
2.3
−71
pOΔVβ8.1CAT
CAT
SV40En
59
100
CACCC box

*Figure 7* Transcriptional activity of the human Vβ8.1 promoter. The constructs of pVβ 8.1CAT, pΔVβ8.1CAT and pOΔVβ8.1CAT have been described in *Materials and Methods*. The transcription start site of the SV40 promoter is indicated by an arrow. (a). Hela cells were transfected with 2 μg of various Vβ8.1 promoter constructs and 0.5 μg of CMV-lac Z using the liposome method. (b). Jurkat cells were transfected with 5 μg of various Vβ8.1 promoter constructs and 1 μg of CMV-lac Z using the DEAE-dextran method. In both cases, the amount of cell extracts used in the CAT assay were optimized according to the β-galactosidase activities.




| | | |
|---|---|---|
| Reporter plasmid (2μg) | pVβ8.1CAT | pVβ8.1CAT |
| Expression plasmid (2μg) | pSG5 | pSG5-htβ |
| Relative CAT activity | 1 | 1.9 |

*Figure 8* Activity of htβ protein on the human Vβ8.1 promoter. 2 μg of pVβ8.1CAT reporter plasmids were cotransfected with 2 μg of expression plasmids pSG5 (Lane 1) or pSG5-htβ (lane 2) into Hela cells using the liposome method (15). Internal control of CMV-lacZ (0.5 μg) was included in each transfection. The relative CAT activities were calculated.

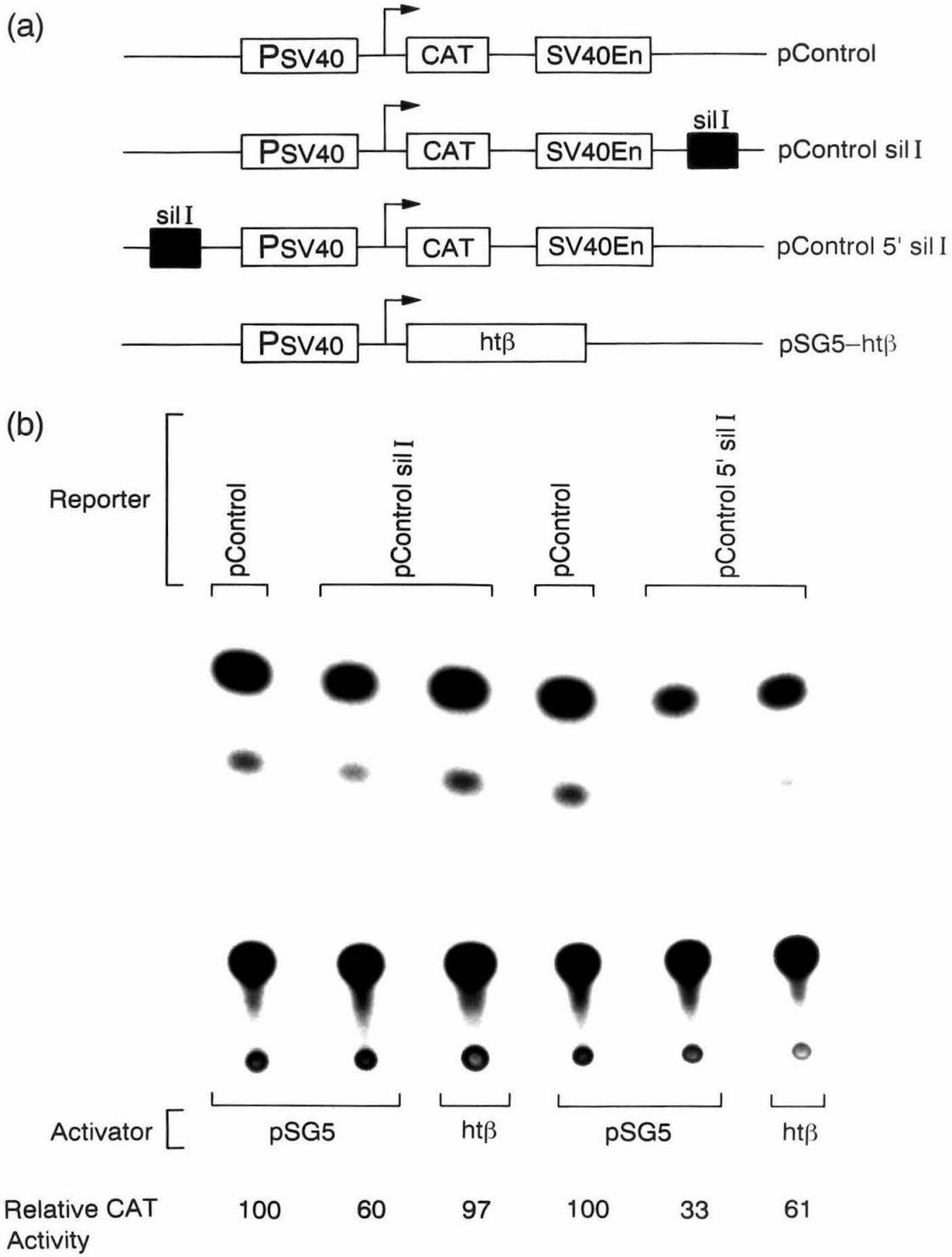

(a)
PSV40
CAT
SV40En
pControl
sil I
pControl sil I
pControl 5' sil I
htβ
pSG5–htβ
(b)
Reporter
pControl
pControl sil I
pControl
pControl 5' sil I
Activator
pSG5
htβ
pSG5
htβ
Relative CAT Activity
100
60
97
100
33
61

*Figure 9* Activity of htβ protein on the mouse TCR α gene silencer. (a). Schematic organization of reporter and expression plasmids. The mouse TCR α gene silencer I was inserted upstream of the SV40 promoter or downstream of the SV40 enhancer. The full length htβ cDNA was inserted at the EcoRI site of the pSG5 vector, which has a SV40 early promoter. The transcription start site of the SV40 promoter is indicated by an arrow. (b). Effect of htβ protein on TCR α silencer I. Hela cells were cotransfected with a reporter plasmid (2 μg) and an activator plasmid pSG5-htβ (2 μg) or the parental vector pSG5, together with 0.5 μg of the internal control plasmid CMV-lacZ. The CAT activity of pControl transfection is considered 100%, and the relative CAT activities for other transfections were calculated.

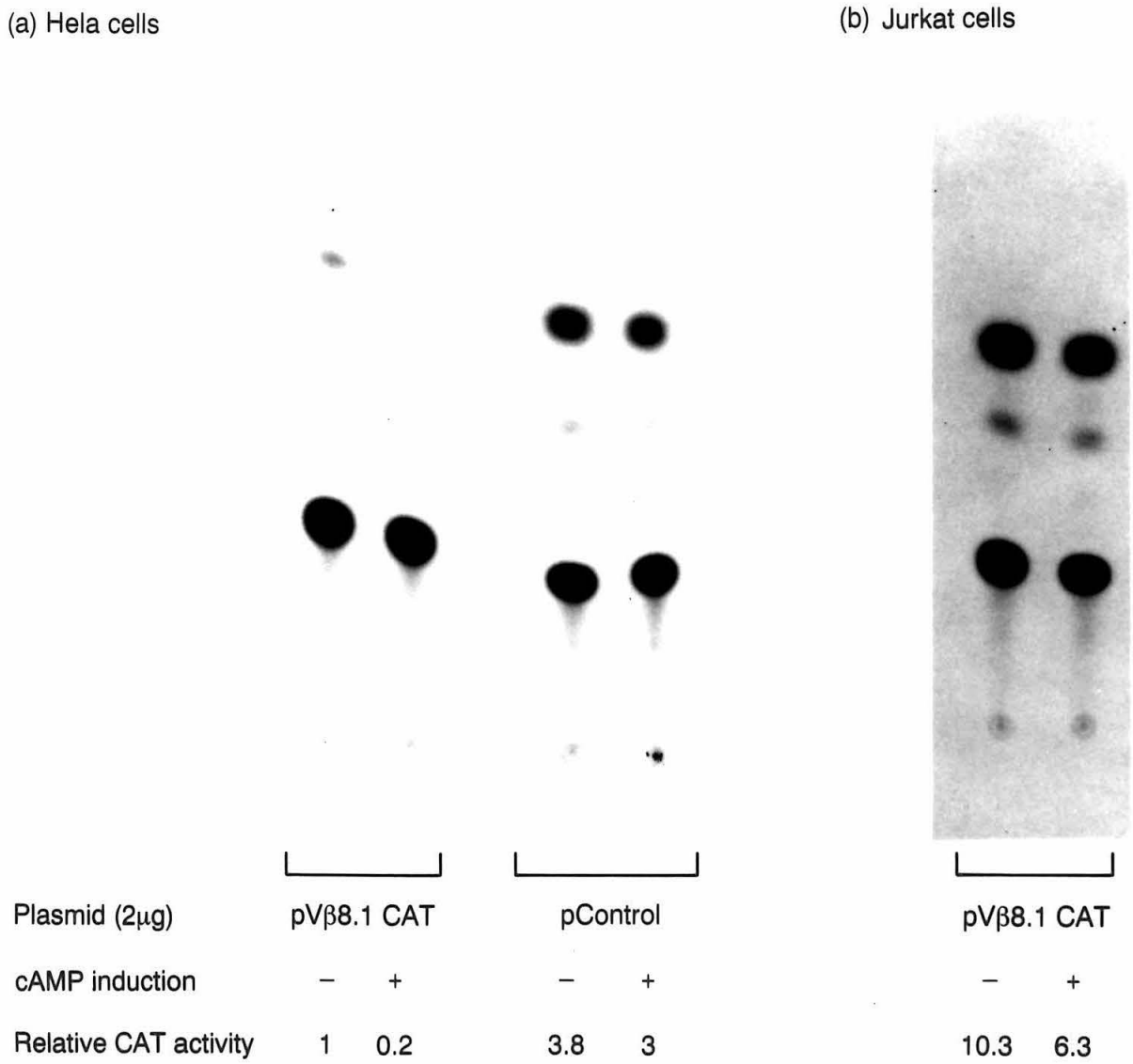

(a) Hela cells
(b) Jurkat cells
Plasmid (2μg)
pVβ8.1 CAT
pControl
pVβ8.1 CAT
cAMP induction
–
+
–
+
–
+
Relative CAT activity
1
0.2
3.8
3
10.3
6.3

*Figure 10* Transcriptional activity of human Vβ8.1 promoter upon cAMP induction. The constructs of pVβ8.1CAT and pControl have been described in *Materials and Methods*. (a). Hela cells were transfected with 2 μg of pVβ8.1CAT or pControl and 0.5 μg of CMV-lac Z using the liposome method with or without cAMP induction. (b). Jurkat cells were transfected with 5 μg of pVβ8.1CAT and 1 μg of CMV-lac Z using the DEAE-dextran method with or without cAMP induction. In both cases, the amount of cell extracts used in the CAT assay were optimized according to β-galactosidase activities.

| Vβ8 gene family member | GGGTG box in promoter sequence |
|---|---|
| Human Vβ8.1 | T G G G G G T G G |
| Human Vβ8.2 | T G G G G G T G G |
| Human Vβ8.3 | T C T G G G T G A |
| Human Vβ8.4 | T G G A G G T G G |
| Human Vβ8.5 | T G G G G G T G G |
| Mouse Vβ8.1 | A A G T G G T G G |
| Mouse Vβ8.2 | G A A G G G T G T |
| Mouse Vβ8.3 | A A G G G G T G T |

*Figure 11* GGGTG boxes found in both human and mouse TCR Vβ8 gene promoters. The GGGTG sequences are boxed, and the flanking sequences also are shown. Sequence are presented in the 5'→3' transcriptional orientation of the gene.

**GGGTG or CACCC box in TCR gene regulatory sequences**

| | | |
|---|---|---|
| Mouse TCR gene α silencer I | G G G T G | (4) |
| | C A C C C | |
| Mouse TCR gene α silencer II | C A C C C | |
| Human TCR β enhancer | G G G T G | |
| Mouse TCR β enhancer | G G G T G | (3) |
| Human TCR α enhancer | C A C C C | |
| Mouse TCR α enhancer | C A C C C | |

*Figure 12* List of GGGTG or CACCC boxes found in TCR gene regulatory sequences. Sequences are presented in the 5'→3' transcriptional orientation of the gene. If more than one copy was found, the copy number is indicated in the bracket.

**GGGTG or CACCC box in non-TCR gene regulatory sequences**

| | |
|---|---|
| Various globin gene promoters | C A C C C |
| GATA-1 promoter | C A C C C |
| Troponin C gene enhancer | G G G T G |
| $H^+/K^+$-ATPase β gene promoter | C A C C C |
| Porpholignogen Deaminase promoter | C A C C C |
| Glucocorticoid receptor promoter | C A C C C |

*Figure 13* GGGTG or CACCC box found in gene regulatory sequences other than TCR genes. Sequences are presented in the 5'→3' transcriptional orientation of the gene.

## Chapter 3

Isolation of a Mouse cDNA That is Homologous to the Human T cell Receptor Promoter Binding Protein Htβ at Zinc Finger Region

**ABSTRACT**

A mouse zinc finger protein (M-zif) has been isolated and characterized. It has four fingers in the zinc finger domain, the putative DNA binding domain. A glutamine-rich region also was found, which may be involved in transcriptional activation. A previously reported cDNA molecule was shown to contain the coding regions of both IL-2Rα receptor and M-zif genes. The results presented here indicate that the cDNA is a chimeric molecule resulting from cloning artifact. The zinc finger domain of M-zif is highly homologous to that of htβ, a human T cell receptor Vβ8.1 gene promoter binding protein. They may have similar DNA binding sites. However, M-zif is not the mouse equivalent of htβ.

## INTRODUCTION

The zinc finger domain is a sequence-specific DNA binding motif that was originally identified in the *Xenopus laevis* transcription factor TFIIIA (1). Since then, many transcription factors with the zinc finger motif have been found in different systems, notably, the yeast ADR1 (2), the *Drosophila* segmentation genes *Kruppel* and *Hunchback* (3-4), and the mammalian Sp1 factor (5-6). DNA sequence analysis of these cloned genes indicates they contain regions that strongly resemble the zinc finger motif of TFIIIA. The typical zinc finger motif consists of approximately 30 amino acids with two cysteine and two histidine residues that form the domain structure by coordinating a $Zn^{2+}$ ion (7). The region of 12 amino acid residues between the Cys-His pair usually are basic and hydrophobic, and may be responsible for DNA sequence-specific binding (8). Another remarkable feature about the zinc finger motif is that there is a region of sequence between the last histidine of one finger and the first cysteine of the next, which is very conservative among the zinc finger-containing proteins of all species (7). Zinc finger proteins have a variety of functions. Yeast ADR1 is a transcription factor required for the transcriptional activation of the alcohol dehydrogenase gene (2), while Sp1 is a general transcription factor that controls many cellular and viral genes (8). The *Drosophila* segmentation genes *Kruppel* and *Hunchback* are involved in control of development in *Drosophila* (3-4), the human zinc finger protein ZFY, located on the Y chromosome, is the major sex-determining gene (9), GL1 (10) and Wilm's tumor genes (11) are proto-oncogenes. Other zinc finger protein, for example, EGR1 and EGR2 are induced by differentiation signals (12-15).

The basic structure of a zinc finger protein consists of an antiparellel β-sheet followed by a helix with a $Zn^{2+}$ ion coordinated to the two cysteines in the loop and to the two histidines (16-18). Zinc finger proteins are able to bind DNA sequence specifically. Both the zinc-finger region and zinc are necessary for binding. In Chapter

1, a human T cell receptor (TCR) Vβ8.1 gene promoter binding protein htβ has been isolated and characterized. Further study (Chapter 2) identified htβ as a CACCC binding protein, which is essential for human TCR Vβ8.1 gene expression. The results from a GenBank search with the htβ gene as a query sequence indicated that htβ is homologous to the noncoding strand of one of the mouse interleukin-2 receptor α chain (IL-2Rα) cDNAs, IL-2R cDNA-J (19). The noncoding strand of the 5' untranslated region of the mouse IL-2Rα has the potential of encoding a mouse zinc finger protein (M-zif). IL-2Rα and M-zif mRNA are transcribed from different DNA strands and transcription proceeds at opposite directions (fig. 1). The M-zif gene has been cloned and analyzed in terms of its interesting relationship with IL-2R α gene. The sequence analysis of M-zif, its comparison with htβ and its functional implication also are reported here.

## MATERIAL AND METHODS

### *Oligonucleotides*

The oligonucletides used for polymerase chain reaction (PCR) of the M-zif gene are listed here.

| | |
|---|---|
| Primer 1 | 5' TGAGACAACACAGCCAGGTTACCC 3' |
| Primer 2 | 5' CCCACCACCACCATCACCACCACC 3' |
| Primer 3 | 5' TGCCCGACAACTGGGTACTAT 3' |
| Primer 4 | 5' CCCAACATCAGCAAGCGTGGC 3' |

The location of all the primers are shown (fig. 1)

### *RT-PCR (reverse transcription of RNA followed by the polymerase chain reaction)*

Poly $(A)^+$ RNAs were prepared as described in Chapter 1. Mouse T cell line EL-4 poly $(A)^+$ RNA (0.5 μg) was used for reverse transcription using cDNA SYNTHESIS SYSTEM PLUS (Amersham). Primer 1 (0.25 μM) was used for priming the first strand cDNA, then 5 μl of first strand cDNA primed by primer 1 was used for PCR with either primer 1 and 2, or primer 1 and 3, or primer 1 and 4. The PCR conditions are as follows: 92°C, 45 sec; 60°C, 60 sec; 72°, 90 sec, for 30 cycles. Finally the PCR products were resolved on an 1% agarose gel. The DNA fragments were isolated and subcloned into the pUC19 vector for sequence analysis (20).

*Genomic DNA PCR*

Mouse BALB/c liver genomic DNA (Clontech) (0.1 μg) was used for PCR with either primer 1 and 3, or primer 1 and 4. The PCR was performed at: 92°C, 45 sec; 55 °C, 90 sec; 72°C, 180 sec, for 30 cycles. The PCR products were resolved on an 1% agarose gel for Southern blot analysis.

*Library construction and screening*

Mouse T cell line EL-4 poly $(A)^+$ RNA (2.5 μg) was used for cDNA synthesis by following the procedure of cDNA SYNTHESIS SYSTEM PLUS (Amersham). Oligo-dT was used for priming the first strand cDNA. The cDNA was then ligated to the λgt11 vector (Stratagene), and packaged using the In Vitro Packaging System for Lambda DNA (Amersham). The EL-4 cDNA library (λgt11) was screened using a 0.75 kb DNA fragment as a probe, which is amplified from EL-4 poly $(A)^+$ RNA with Primer 1 and 2 by the PCR method. The screening was performed according to the manufacturer's protocol. The EcoRI inserts from positive plaques were subcloned into pUC19 for sequence analysis using the dideoxy chain termination method for both strands.

*Southern and Northern analysis*

Mouse BALB/c liver genomic DNA (5 μg) was used for Southern blot analysis. Preparation of poly $(A)^+$, Southern and Northern blot analyses were done as described in Chapter 1. Both Jurkat human T cells and EL-4 mouse T cells were cultured at 37°C, and the poly $(A)^+$RNAs from both cell lines were isolated.

## RESULTS

*Isolation of the M-zif gene*

To analyze the relationship between IL-2Rα and M-zif genes, and whether the M-zif gene is actively transcribed, the RT-PCR cloning method was used for isolating the M-zif gene. Primer 1 that complements the M-zif gene coding strand was used for priming first strand cDNA synthesis, and PCR amplification of the first strand cDNA with Primer 1 and 2 was performed. A 0.75 kb DNA fragment was isolated from Primer 1 and 2 amplified PCR products. Sequence analysis of this 0.75 kb fragment verified that it is part of the M-zif gene. A EL-4 cDNA library (λgt11) was constructed, and $10^6$ plaques were screened with the M-zif 0.75 kb fragment as a probe. Two positive plaques were found. The two M-zif cDNAs are 0.65 kb (M-zif 2) and 1.3 kb (M-zif 7) in length. M-zif 2 and M-zif 7 overlap by approximately 0.65 kb in the 5' region (fig. 2).

*DNA sequence of M-zif*

The DNA sequence of M-zif was determined (fig. 3). Sequence analysis of the M-zif gene reveals a long open reading frame of 732 nucleotides which is terminated with two in-frame stop codons. Since the open reading frame continues to the 5' end of the cDNA for both M-zif 2 and M-zif 7, very likely the full length cDNA has not been

isolated yet. The partial M-zif cDNA encodes a protein of 244 amino acids. Analysis of the M-zif amino acid sequence translated from the cDNAs reveals a few interesting features. Firstly, there is a zinc finger domain at the 3' end of the open reading frame. The first two fingers have the typical finger motif of Cys-$X_2$-Cys-$X_3$-Phe-$X_5$-Leu-$X_2$-His -$X_3$-His. The third and fourth fingers have the modified forms of Cys-$X_2$-Cys-$X_3$-Phe-X8-His-$X_3$-His and Cys-$X_2$-Cys-$X_3$-Phe-$X_5$-Leu-$X_2$-His-$X_3$-Cys respectively. The results from a GenBank search indicate that the M-zif gene encodes a novel zinc finger protein. Secondly, there is a glutamine-rich region (residue 17-25) in the 5' region of the M-zif gene. Finally, a segment rich in basic amino acids is found upstream of the zinc finger domain( residues 53-70)

*Southern blot and Northern blot analysis*

To determine whether the M-zif gene has one copy or multiple copies, BALB/c mouse liver DNA was digested with four restriction enzymes for Southern blot analysis (fig. 4). A single band was seen with each restriction enzyme digest. M-zif gene expression was studied in the EL-4 murine T cell line and Jurkat human T cell line with a $^{32}$P-labeled M-zif 2 fragment. A 5 kb band was detected in both RNA samples (fig. 5).

*Genomic DNA PCR and RT-PCR analysis of M-zif gene*

Attempts were made to amplify the mouse genomic DNA fragment that includes the protein coding regions of both M-zif and the mouse IL-2 receptor $\alpha$ gene using the PCR method. Primer 1 and 3, or Primer 1 and 4 were used for PCR amplification of mouse genomic DNA (see figure 1). RT-PCR also was performed. The first strand cDNA was synthesized by using Primer 1, then the cDNA was subject to PCR amplification with Primer 1 and 2, or Primer 1 and 3, or Primer 1 and 4. The PCR products from both genomic DNA and RT-PCR were used for Southern blot analysis with M-zif 2 cDNA as a probe. There was a 0.75 kb band detected with RT-PCR

products synthesized with Primer 1 and 2, which was described earlier, but no bands can be detected from either genomic DNA PCR or RT-PCR products amplified with either Primer 1 and 3 or Primer 1 and 4. (fig. 6).

## DISCUSSION

In Chapter 1 and 2, a novel zinc finger protein htβ, which binds to the human T cell receptor Vβ8.1 gene promoter, has been isolated and functionally characterized. GenBank search with htβ as a query sequence identified that it is homologous to the noncoding strand of one of the mouse IL-2 receptor gene α (IL-2Rα) cDNAs, IL-2R cDNA-J (19). The 5' untranslated region of the IL-2R α gene encodes a mouse zinc finger protein (M-zif), but the IL-2R α gene and M-zif genes are transcribed in opposite orientation (fig. 1). To determine if the M-zif gene is located next to the IL-2R α gene on the chromosome and whether it is the mouse counterpart of htβ, molecular cloning of the M-zif gene was necessary. M-zif was cloned by using the polymerase chain reaction method. Two M-zif cDNA clones were isolated by screening a cDNA library made from mouse EL-4 cell derived mRNA probed with a M-zif gene PCR fragment. The sequence analysis of the M-zif gene indicated it has a zinc finger domain with four fingers in it. The first two finger motifs have the general form, while the third and fourth fingers have been modified. Especially the fourth one, instead of having Cys-Cys and His-His pairs, it has Cys-Cys and His-Cys pairs. The $C_2$-H-C type of finger has been found in the human IFN-β gene promoter binding protein PRDII-BF1 (21) and the sterol regulatory element binding protein (22). The $C_2$-$H_2$ type of finger, which was found in the first three fingers of M-zif, was shown to be sufficient for specific DNA binding of the transcription factor TFIIIA (1) and Sp1(8). The $C_2$-H-C finger, which is the fourth finger of M-zif, is required for binding to single-stranded

DNA or genomic RNA of retroviruses (23-24). In the case of PRDII-BF1 (21), the $C_2$-H-C finger is not required for human IFN-β gene promoter binding. The function of the $C_2$-H-C finger in the M-zif gene is yet to be identified. The comparison of finger regions between htβ and M-zif indicate they are 75% homologous (fig. 7). Surprisingly the two finger domains are almost identical at those amino acid residues which presumably interact with DNA (boxed amino acids in fig. 7) according to the crystal structure analysis of a mouse immediate early protein Zif268 (25). Therefore, possibly M-zif recognizes the same or very similar DNA sequence as htβ. Besides the zinc finger region, M-zif has a short glutamine-rich region that may be involved in transcriptional activation. The glutamine-rich region has been characterized as a transcriptional activation domain in the human Sp1 transcription factor (8). A highly positively-charged region also is found, which may be the putative region responsible for nuclear localization of M-zif (26). Once the full length cDNA of M-zif has been isolated, further sequence analysis can be pursued.

The possibility that the IL-2R α and M-zif genes are located at the same location on the chromosome is intriguing because the coding region of M-zif overlaps with the 5' untranslated region of IL-2Rα. The expression of one gene may have a direct effect on the expression of the other gene because of structural changes of that region of chromatin. Southern blot analysis of mouse DNA with a M-zif probe revealed a single band in mouse DNA digests with either EcoRI (5 kb) or HindIII (7.4 kb) or BamHI (1.35 kb) or XbaI (6.8 kb). The Southern blot results with an IL-2Rα probe revealed multiple bands for each of five restriction enzyme digestions (27). A band approximately 5 kb can be seen on both blots with EcoRI digestion. From Southern blot results, the question of whether M-zif and IL-2Rα are located adjacent but in opposite orientation to each other cannot be resolved. Northern blot analysis of the murine EL-4 cell line with a M-zif probe reveals a single band, approximately 5 kb. However, the Northern blot results of the same cell line with an IL-2Rα probe that is located in the coding

region of IL-2Rα indicated a 4.1 kb band only. At least two bands on the Northern blot detected with a M-zif probe were expected, representing the two transcripts from the IL-2Rα promoter and the M-zif promoter. At least one band was expected to be observed on Northern blot analysis with IL-2Rα probe, which represents the transcript from IL-2Rα promoter. In short, a common band that represents the transcript from the IL-2Rα promoter is supposed to be detected on Northern blot analysis with the M-zif probe and the IL-2Rα probe if the IL-2R cDNA-J is not a chimeric molecule. Northern blot analysis of the EL-4 cell line with both M-zif and IL-2Rα probes revealed no common band, suggesting the IL-2R cDNA-J is probably a chimeric molecule that combined the M-zif and IL-2Rα cDNA molecules. In addition, when two sets of primer pairs, which complement the protein coding regions of either the IL-2R α gene or the M-zif gene, were used for mouse genomic DNA PCR and RT-PCR amplification, there were no bands seen on a Southern blot detected with the M-zif 2 probe. An expected band was observed on the control lane in which the PCR product amplified with two primers located in the M-zif coding region was loaded. On the basis of these data, very likely the IL-2R cDNA-J described by Froussard et al. (19) was a cloning artifact (20). The chimeric DNA molecule has one cDNA from the M-zif gene and another from the IL-2R α gene. As a matter of fact, the authors also raise the possibility that the IL-2R cDNA-J represents a cloning artifact. The comparison of the IL-2R cDNA-J sequence with the murine p55 IL-2R cDNA (28) and the human IL-2R gene (29-30) show they diverge extensively in the 5' untranslated region. Although the IL-2R cDNA-J is very likely a chimeric cDNA molecule, some genes may truly overlap with each other because of their related functions. The transcription of one gene may change the chromatin structure and make the other gene accessible to the transcription machinery, causing the activation or repression of that gene.

Since M-zif and htβ sequences have almost no homology except for the finger domain, very likely M-zif is not the htβ equivalent in mouse. Northern blot analysis of

both the murine EL-4 and the human Jurkat cell lines with a M-zif probe indicates a single band of approximately 5 kb. Northern blot analysis of the Jurkat cell line with a htβ probe reveals three bands, approximately 4.2, 7.6 and 8.6 kb (Chapter 1). Both sequence and Northern blot results clearly demonstrate that M-zif is not the htβ counterpart in mouse. Although htβ and M-zif are different genes and may carry out different gene regulating functions, they possibly recognize very similar or identical DNA sequences because of their highly homologous zinc finger domains. This may represent a mechanism of how zinc finger proteins regulate gene expression positively or negatively. This speculation is based on the fact that htβ and M-zif have almost identical amino acid residues that are currently known to be important for sequence-specific binding. The possibility that other amino acid residues in the finger or out of the finger region also contribute to sequence-specific recognition by both htβ and M-zif cannot be ruled out. Comparing the binding of htβ and M-zif proteins with their putative DNA binding sites will be useful for solving the problem.

Finally, many zinc finger proteins have been characterized since the zinc finger motif first was described in TFIIIA. These proteins have been shown to have both sequence-specific binding and gene regulation functions. How the amino acids in the finger interact with nucleotides in the DNA sequence is still not very clear. There is speculation that a code system may dictate the interaction between amino acid residues and nucleotides (31). Such a system would predict which amino acid residue can interact with which nucleotide. Accumulation of additional data on zinc finger protein/DNA interaction will facilitate the derivation of rules for such a code.

REFERENCES

1. Miller, J., McLachlan, A.D. and Klug, A. (1985) EMBO J. 4:1609-1614.

2. Hartshorne, T.A., Blumberg, H. and Young, E.T. (1986) Nature 320:283-287.

3. Rosenberg, U.B., Schroder, C., Preiss, A., et al. (1986) Nature 319:336-339.

4. Tauz, D., Lehman, R., Schnurch, H., et al. (1987) Nature 327:383-389.

5. Kadonaga, J.T., Carner, K.R., Masiarz, F.R. and Tjian, R. (1987) Cell 51:1079-1090.

6. Kadonaga, J.T., Courey, A.J., Ladika, J. and Tjian, R. (1988) Science 242:1566-1570.

7. Evans, R.M. and Hollenberg, S.M. (1988) Cell 52:1-3.

8. Mitchell, P. and Tjian, R. (1989) Science 245:371-378.

9. Page, D.C., Mosher, R., Simpson, E.M., et al. (1987) Cell 51:1091-1104.

10. Kinzler, K.W., Ruppert, J.M., Bigner, S.H. and Vogelstein, B. (1988) Nature 332:371-374.

11. Call, K.M., Glaser, T., Ito, C.Y., et al. (1990) Cell 60:509-520.

12. Sukhatme, V.P., Cao, X.M., Chang, L.C., et al. (1988) Cell 53:37-43.

13. Christy, B.A., Lau, L.F. and Nathans, D. (1988) Proc. Natl. Acad. Sci. USA 85:7857-7861.

14. Joseph, L.J., Lebeau, M.M., Jamieson, G.A., et al. (1988) Proc. Natl. Acad. Sci. USA 85:7264-7168.

15. Chavrier, P., Zerial, M., Lemaire, P., Almendral, J., Bravo, R. and Charnay, P. (1988) EMB0 J. 7:29-35.

16. Parraga, G., Horvath, S.J., Eisen, A., et al (1989) Science 245:1489-1492.

17. Lee, M.S., Gippert, G.P., Soman, K.W., Case, D.A. and Wright, P.E. (1989) 245:635-637.

18. Berg, J.M. (1990) Annu. Rev. Biophys. Biophys. Chem. 19:405-421.

19. Froussard, P., Chastagner, P., Somme, G., et al (1988) J. Immunol. 14:1358-1364.

20. Manaitis, T., Fritsch, E.F. and Sambrook, J. (1989) Molecular Cloning, A Laboratory

Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

21. Fan, C.-M. and Maniatis, T. (1990) Genes Dev. 4:29-42.

22. Rajavashisth, T.B., Taylor, A.K., Andalibi, A., Sevenson, K.L. and Lusis, A.J. (1989) 245:640-643

23. Covey, S.N. (1986) Nucleic Acids Res.14:623-633.

24. Berg, M.J. (1986) Science 232:485-487.

25. Pavletich, N.P. and Pabo, C.O. (1991) Science 252:809-817.

26. Dingwall, C. and Laskey, R.A. (1986) Ann. Rev. Cell Biol. 2:367-390.

27. Miller, J., Malek, T.R., Leonard, W.T., Greene, W.C., Shevath, E.M., and Germain,R.N. (1985) J. Immunol. 134:4212-4217.

28. Shimuza, A., Kondo, S., Takeda, S., (1985) Nucleic Acids Res. 13:1505-1516.

29. Ishida, N., Kanomori, H., Noma, T., et al. (1985) Nucleic Acids Res. 13:7580-7589.

30. Suzuki, N., Matsunami, N., Kanamori, H. (1987) J. Biol. Chem. 262:5079-5086.

31. Thiesen, H.-J. and Bach, C. (1991) FEBS Lett. 283:23-26.

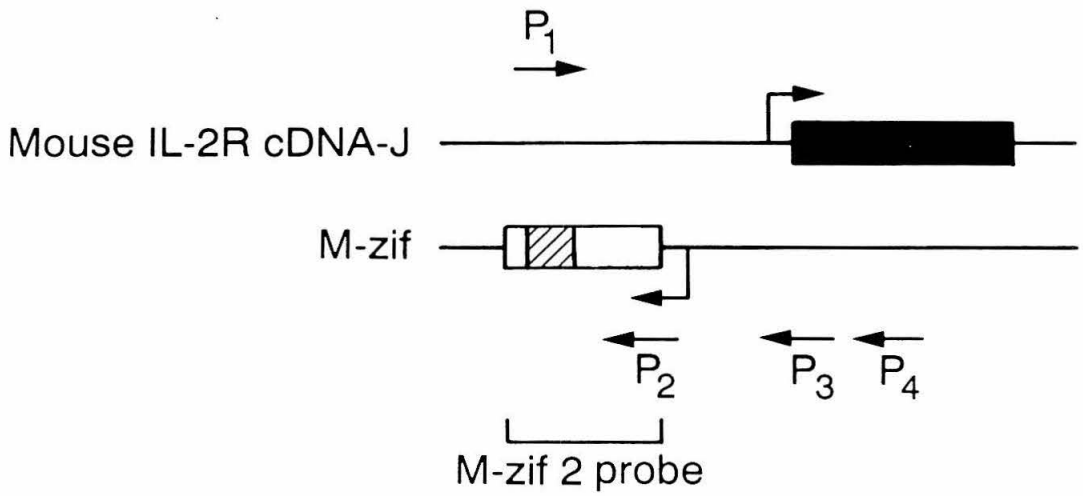

P1
Mouse IL-2R cDNA-J
M-zif
P2
P3
P4
M-zif 2 probe

*Figure 1* The IL-2 receptor (IL-2R) α gene and the M-zif gene are contained in the same cDNA molecule, the IL-2R cDNA-J (20). In this diagram, the two genes are transcribed in opposite directions, indicated by the arrows. The locations of the arrows do not indicate the actual transcriptional start sites since they are not known for both genes. The IL-2Rα and M-zif coding regions are represented by black and blank boxes respectively, and the region with hatched lines indicates the zinc finger domain of M-zif gene. $P_1$, $P_2$, $P_3$ and $P_4$ are the primers used for polymerase chain reaction. The locations and directions of these primers are shown in the figure. The location of M-zif 2 probe also is indicated.

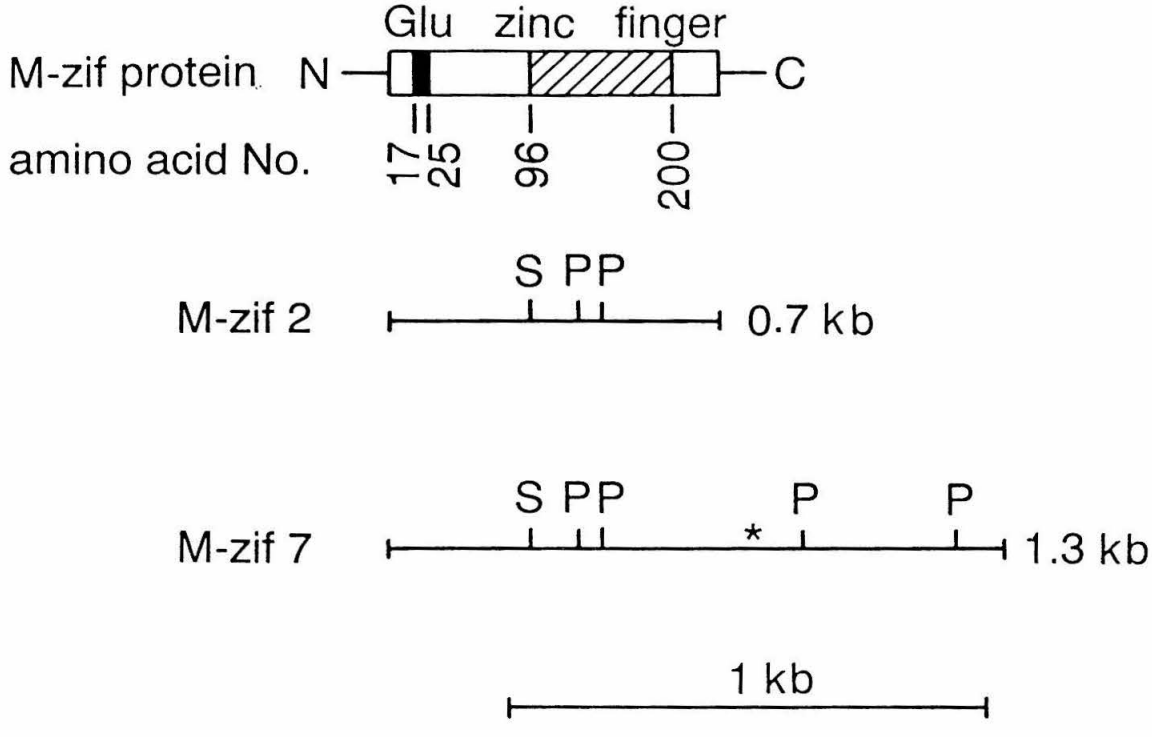

Glu zinc finger
M-zif protein N C
amino acid No.
17
25
96
200
S PP
M-zif 2
0.7 kb
S PP P P
M-zif 7
*
1.3 kb
1 kb

*Figure 2* cDNAs and predicted protein structure of M-zif. The overlapping cDNAs M-zif 2 and M-zif 7 were isolated from an EL-4 T cell cDNA library. Restriction enzyme sites: P, Pst I; S, Sac I. A schematic diagram of the predicted the M-zif protein structure is shown, with 2 cDNA clones aligned to the protein structure. The interesting motifs in the protein are indicated: Glu, glutamine-rich region; zinc finger, zinc finger domain.

```
   1  CGGGAAGGGGGGAGCCACGGCGTCATCCAAGACCTCAGTCTTCTGCACCAGCATGCCCAG   60
   1  R  E  G  G  S  H  G  V  I  Q  D  L  S  L  L  H  Q  H  A  Q     20

  61  CAGCAACCGGCCCAGCACCACCGCGATGTATTACTGAGCAGCGGTAGCAGGACTGATGAG  120
  21  Q  Q  P  A  Q  H  H  R  D  V  L  L  S  S  G  S  R  T  D  E     40

 121  CACGGCAACCAGGAGCCAAAGCAGGACGCTAACGTCAAAAAGGCAAAGAGGCCAAAGCCA  180
  41  H  G  N  Q  E  P  K  Q  D  A  N  V  K  K  A  K  R  P  K  P     60

 181  GAATCTCAGGGAATCAAAGCCAAGAGGAAGCCAAGTGCATCTTCCAAACCTTTGGTTGGA  240
  61  E  S  Q  G  I  K  A  K  R  K  P  S  A  S  S  K  P  L  V  G     80

 241  GAGGGAGAAGGTGCCGTCCTGTCCCCAAGTCAGAAACCTCATATCTGTGATCACTGTAGT  300
  81  E  G  E  G  A  V  L  S  P  S  Q  K  P  H  I  C  D  H  C  S    100

 301  GCTGCTTTCAGGAGCTCCTACCACCTGCGGAGACATGTCCTCATCCACACGGGCGAAAGG  360
 101  A  A  F  R  S  S  Y  H  L  R  R  H  V  L  I  H  T  G  E  R    120

 361  CCTTTCCAGTGCAGCCAGTGTAGTATGGGTTTCGTCCAGAAATACCTACTGCAGAGACAC  420
 121  P  F  Q  C  S  Q  C  S  M  G  F  V  Q  K  Y  L  L  Q  R  H    140

 421  GAGAAAATTCACAGCAGAGAGAAACCCTTTGGGTGCGATCAGTGCAGCATGAAGTTTATT  480
 141  E  K  I  H  S  R  E  K  P  F  G  C  D  Q  C  S  M  K  F  I    160

 481  CAGAAGTACCATATGGAGAGACACAAGAGGACACATAGTGGAGAAAAGCCATACAAGTGT  540
 161  Q  K  Y  H  M  E  R  H  K  R  T  H  S  G  E  K  P  Y  K  C    180

 541  GACACTTGCCAACAGTATTTTTCAAGGACTGACAGATTGTTGAAGCACAGGCGCACGTGT  600
 181  D  T  C  Q  Q  Y  F  S  R  T  D  R  L  L  K  H  R  R  T  C    200

 601  GGTGAAGCCATAGCAAAAGGAGCCGCTAGTGCAGAACCTGGGTCATCAAACCATAACAGT  660
 201  G  E  A  I  A  K  G  A  A  S  A  E  P  G  S  S  N  H  N  S    220

 661  ATGGGTAACCTGGCTGTGTTGTCTCAGGGAAATACAAGTTCCTCAAGGAGAAAGTCGAAT  720
 221  M  G  N  L  A  V  L  S  Q  G  N  T  S  S  S  R  R  K  S  N    240

 721  CGAAAACATAGCTAGTGAAAATAAGGAACACAAGACTGGCAAAACAAATGAATCACAAAT  780
 241  R  K  H  S  *  *                                              260

 781  GTCAAACAATATAAACATGCAGAGTTATTCAGTAGAAATGCCTACTGTGTCTACCAGTGG  840

 841  GAGCATAATTGGCACTGGAATAGATGAACTGCAGAAAAGGGTGCCAAAATTGATCTTTAA  900

 901  GAAAGGAAGCAGAAAGAATGCAGATAAAAGCTACCTTAATTTTGTGTCACCGTTACCAGA  960

 961  CGTAGTTGGGCAGAAATCCTTGTCTGGTAAGCCAGGTGGCTCACTTGGCATCGTATCGAA 1020

1021  TAATAGTGTGGAGACCATTAGTCTTCTCCAAAGTACAAGTGGCAAGCAAGGCCCGATAAG 1080

1081  TAGTAATTATGATGATGCCATGCAGTTTTCAAAGAAGAGAAGATACCTACCAACTGCCAG 1140

1141  CAGCAACAGTGCCTTTTCTATCAATGTAGGACACATGGTCTCCCAGCAGTCAGTCATTCA 1200

1201  GTCTGCAGGGGTCAGTGTTTTGGACAATGAGGCGCCATTGTCACTTATTGACTCCTCAGC 1260

1261  TCTAAATGCTGAAATTAAGTCTTGTCACGACAAGTCTGGAATTC  1304
```

*Figure 3* DNA sequence and predicted amino acid sequence of M-zif. The amino acid sequences predicated from the longest open reading frame of M-zif are shown. The zinc finger motifs are shown with an underline; the glutamine-rich region is indicated by a double line and glutamines are circled. The basic segment which is located before the zinc finger is shown with broken lines. The two stop codons in the longest open reading frame are indicated with "*".

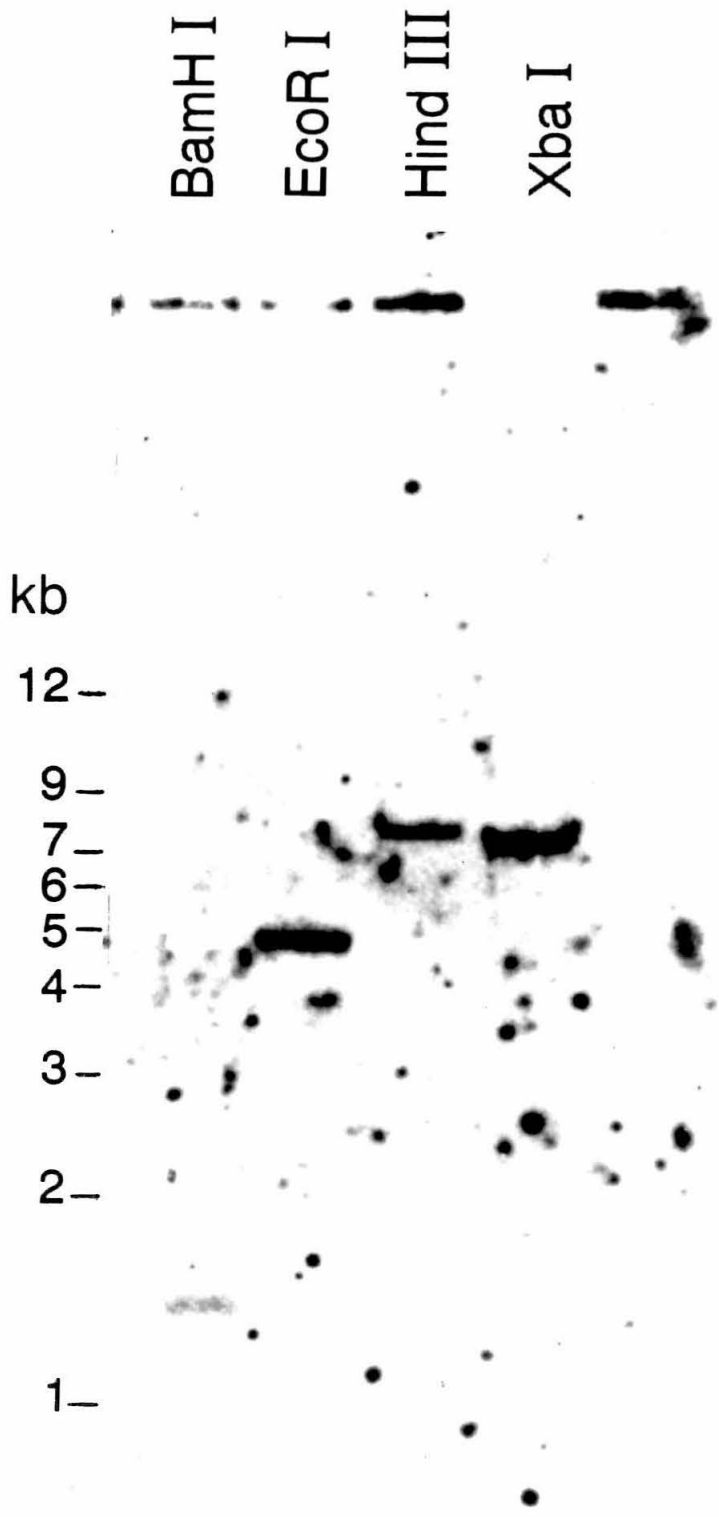

BamH I
EcoR I
Hind III
Xba I
kb
12
9
7
6
5
4
3
2
1

*Figure 4* Southern blot analysis of mouse liver genomic DNA probed with the M-zif 2 clone. The genomic DNA was digested with various restriction enzymes indicated in the figure.

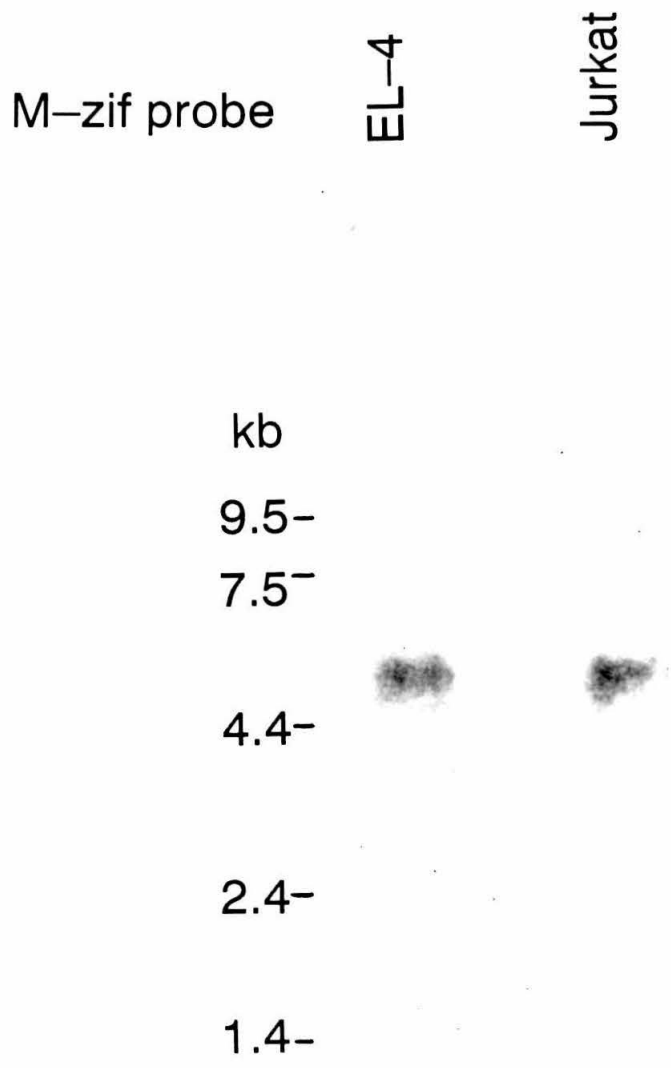

M–zif probe
EL–4
Jurkat
kb
9.5-
7.5-
4.4-
2.4-
1.4-

*Figure 5* Northern blot analysis of poly $(A)^+$ RNA from EL-4 and Jurkat cell lines. Poly $(A)^+$ RNA (10 μg) was loaded on each lane. The blot was hybridized with the 0.65 kb insert of M-zif 2.

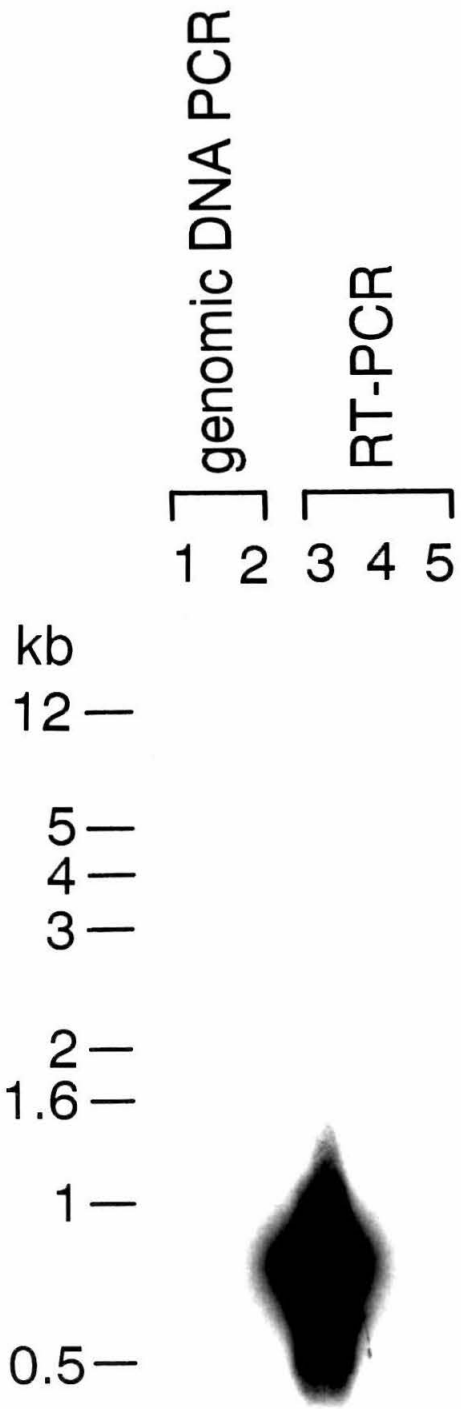

genomic DNA PCR
RT-PCR
1 2 3 4 5
kb
12
5
4
3
2
1.6
1
0.5

*Figure 6* Southern blot analysis of genomic DNA PCR and RT-PCR products with M-zif 2 probe. Lane 1 and 2, genomic DNA PCR with Primer 1 and 3 and Primer 1 and 4 respectively. Lane 3-5, RT-PCR with Primer 1 and 2, Primer 1 and 3 and Primer 1 and 4 respectively.

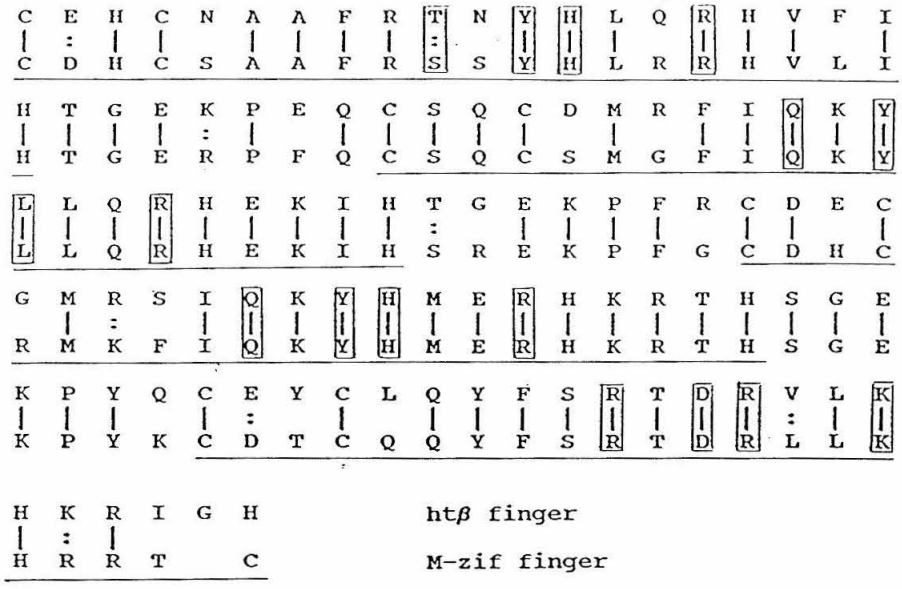

C E H C N A A F R T N Y H L Q R H V F I
C D H C S A A F R S S Y H L R R H V L I
H T G E K P E Q C S Q C D M R F I Q K Y
H T G E R P F Q C S Q C S M G F I Q K Y
L L Q R H E K I H T G E K P F R C D E C
L L Q R H E K I H S R E K P F G C D H C
G M R S I Q K Y H M E R H K R T H S G E
R M K F I Q K Y H M E R H K R T H S G E
K P Y Q C E Y C L Q Y F S R T D R V L K
K P Y K C D T C Q Q Y F S R T D R L L K
H K R I G H
H R R T C
htβ finger
M-zif finger

*Figure 7* Alignment of zinc finger domains between htβ and M-zif. The top line shows the htβ finger region (amino acid residues 173-278), and the bottom line indicates the M-zif finger domain (amino acid residues 96-200). The amino acid residues which are possibly important for interacting with nucleotides in DNA sequence (25) are boxed.

# CONCLUSION

A gene encoding a human zinc finger protein, htβ, has been isolated and characterized. It has a zinc finger domain of $Cys_2$-$X_{12}$-$His_2$ class and an acidic domain. The zinc finger domain is one of the well-known DNA binding domains (1). The amino acid residues that are located between the second conservative cysteine and the first conservative histidine are possibly responsible for the DNA binding function (2). Usually positively charged amino acid residues are found in that region. Each htβ zinc finger has a few positive amino acids, which may interact with guanines of the CACCC box. The acidic region is a putative transcription activation domain. The acidic domain first was described in the yeast transcription factor GAL4 (3). Features of the acidic domain generalized from GAL4 factor are: negative charges and α-helical structure. The acidic domain of htβ has both of these features. Since htβ has the two well-characterized domains for a transcription factor, we believe it is a novel transcription factor involved in gene regulation.

In terms of its function, htβ first was thought to be a T cell-specific transcription factor. However, further study indicated that it is a general transcription factor that bind to the well-known CACCC box. The CACCC box first was shown to be essential for globin gene transcription (4), and later found to be important for the expression of many other genes (5-9). The results in this thesis also indicate that the CACCC box is necessary for efficient transcription of human T cell receptor gene Vβ 8.1. Htβ can activate the transcription of the human TCR Vβ8.1 gene. Since the htβ gene is expressed in Hela cell, B cell, macrophage and αβ T cell lines, very likely htβ functions as a general transcription factor by interacting with the CACCC box in various gene regulatory sequences.

How the htβ protein works as a general transcription factor in not clear. Since ht β has an acidic domain that is even more negatively charged than that of yeast transcription factor GAL4, it will be interesting to determine if htβ interacts directly with the protein TFIID that is a TATA box binding protein. TFIID has been proposed

as the target of the acidic activation region (3), and the interaction between herpes virus protein VP16 that bears an acidic activation region and TFIID has been reported (10). Investigations of the interaction between htβ and TFIID (fig. 1) will be very important to understand how htβ works as a transcription activator. There is a cAMP response element (CRE) just adjacent to the CACCC box in the human Vβ8.1 promoter. Possibly there is a protein-protein interaction between CRE binding protein and htβ. The interaction between the CACCC binding protein and other transcription factors has been reported (9, 11-12). A speculation is that some transcription factors may bind to htβ that bears an acidic activation domain that interacts with TFIID (fig. 2). This may explain how these transcription factors activate gene expression.

Besides binding to the human TCR Vβ8.1 promoter, htβ also binds to the mouse TCR α silencer. The CACCC box has been observed in many TCR β gene promoters (13-14), TCR α and β enhancers (15-18), and the mouse TCR α silencer (19). The CACCC box in the TCR δ enhancer does not have binding activity (20), and there is no CACCC box in the TCR γ enhancer (21). The regulatory function of htβ that binds to the CACCC box in human TCR Vβ8.1 promoter and mouse TCR α silencer has been studied. On the basis of the above facts, we propose that the CACCC box binding protein may act as a major regulator of TCR gene expression. In αβ T cells, the CACCC factor interacts with the CACCC box in TCR α and β enhancers to activate transcription. Also, the CACCC factor binds to the TCR α silencer to relieve its silencing effect on the TCR α enhancer. In γδ T cells, the CACCC factor is not expressed or not functional, resulting in the TCR α silencer to be active, which decreases the enhancing effect of the TCR α enhancer. Hence, the TCR α gene expression is repressed in γδ T cells. Comparison of the htβ gene expression in αβ T cells versus γδ T cells and functional analysis of the CACCC boxes located in TCR α and β enhancers will be very important for revealing the regulatory functions of the CACCC factors on TCR gene expressions in both αβ and γδ T cells.

In addition, a novel gene encoding a murine zinc finger protein (M-zif) has been cloned. M-zif is highly homologous to htβ at the zinc finger domains, although M-zif is not the htβ counterpart in mouse. Since M-zif and htβ have almost identical amino acid residues that are currently known to be important for sequence-specific binding (22), they may recognize very similar or identical DNA sequences. The next goal will be to test the binding of M-zif to the CACCC box. If M-zif can bind to the CACCC box, the regulatory functions carried out by htβ and M-zif will become very interesting. The observation that two different transcription factors can bind to the same site may provide insight into a new mechanism of how zinc finger proteins regulate gene expression. Finally, it also is important to clone the htβ counterpart in mouse since htβ has been shown to regulate the murine TCR α gene silencer. Northern blot analysis of the human Jurkat cell line with a htβ probe did not reveal any band. Using a htβ probe to screen the mouse cDNA library at low stringent conditions may be an appropriate approach to isolate the murine equivalent of htβ.

REFERENCES

1. Miller, J., MaLachlan, A.D. and Klug, A. (1985) EMBO J. 4:1609-1614.

2. Mitchell, P. and Tjian, R (1989) Science 245:371-378.

3. Ptashne, M. (1988) Nature 335:683-689.

4. Dierks, P., van Ooyen, A., Cochran, M.D., Dobkin, C., Reiser, J. and Weissmann, C. (1983) Cell 32:695-706.

5. Tsai, S.-F., Strauss, E. and Orkin, S.H. (1991) Genes Dev. 5:919-931.

6. Parmacek, M.S., Vora, A.J., Shen, T., Barr, E., Jung, F. and Leiden, J.M. (1992) Mol. Cell Biol. 12:1967-1976.

7. Masda, M., Oshuma, K.-I., Tamura, S., et al. (1991) J. Bio. Chem. 266:21584-21588.

8. Govindan, M.V., Pothier, F., Leclerc, S., Palaniswami, R. and Xie, M. (1991) J. Steroid Biochem. Mol. Biol. 40:317-323.

9. Xu, M. and Stavnerzer, J. (1992) EMBO J. 11:145-155.

10. Stinger, K.F., Ingles, C.J. and Greenblatt, J. (1990) Nature 345:783-786.

11. Frampton, J., Walker, M., Plumb, M. and Harrison, P.R. (1990) Mol. Cell Biol. 10:3838-3842.

12. Muller, M., Baniahmed, C., Kaltsschmidt, C. and Renkawith, R. (1991) Mol. Endo. 5:1498-1503.

13. Anderson, S.J., Chou, H.S., and Loh, D.Y. (1988) Proc. Natl. Acad. Sci. USA 85:3551-3554.

14. Siu, G., Strauss, E.C., Lai, E. and Hood, L. (1986) J. Exp. Med. 164:1600-1614.

15. Ho, I.-C., Yang, L.-H., Morle, G. and Leiden, J.M. (1989) Proc. Natl. Acad. Sci. USA 86:6714-6718.

16. Gottschalk, L.R. and Leiden, J.M. (1990) Mol. Cell Biol. 10:5486-5495.

17. Winoto, A. and Baltimore, D. (1989) EMBO J. 7:729-733.

18. Krimpenfor, P., deJong, R., Uematsu, Y., et al. (1988) EMBO J. 7:745-750.

19. Winoto, A. and Baltimore, D. (1989) Cell 59:649-655.

20. Redondo, J.M., Hata, S., Brocklehurst, C. and Krangel, M.S. (1990) Science 247:1225-1229.

21. Kappes, D.J., Browne, C.P. and Tonegawa, S. (1991) Proc. Natl. Acad. Sci. USA 88:2204-2208.

22. Pavletich, N.P. and Pabo, C.O. (1991) Science 252:809-817.

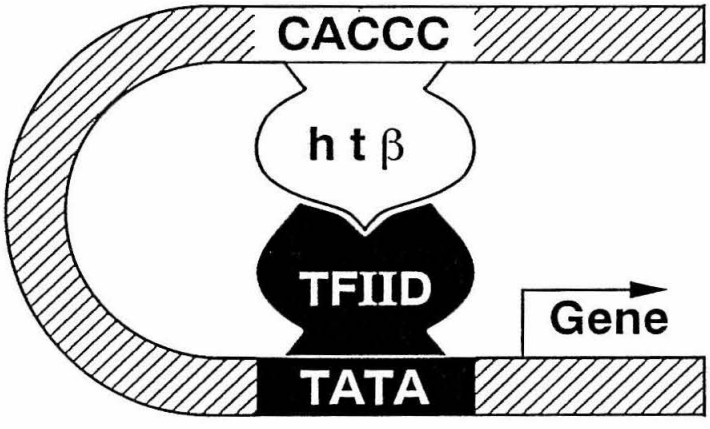

CACCC
h t β
TFIID
Gene
TATA

*Fig. 1* The model of gene activation by the htβ protein. The htβ protein has a zinc finger domain for DNA binding and an acidic activation region. The zinc finger domain binds to the CACCC box and the acidic domain contacts the TFIID that binds at the TATA box. These interactions may loop the intervening DNA since the CACCC box is not adjacent to the TATA box.

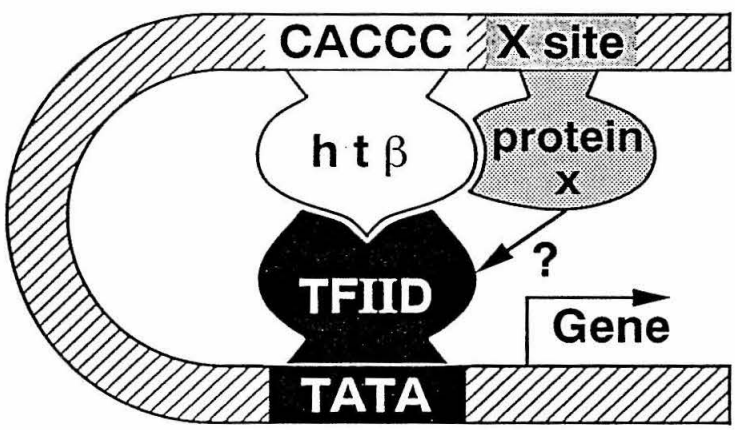

CACCC
X site
h t β
protein
x
?
TFIID
Gene
TATA

*Fig. 2* The model of interaction between the htβ protein and the other transcription factor. The htβ protein binds to the CACCC box and interacts with the TFIID that binds to the TATA box. Another transcription factor, termed protein X, binds to its cognate X site and interacts with htβ protein, then increases or decreases the gene transcriptional activity. Again, these interactions may loop the intervening DNA.