Title
Developmental regulation of lymphoid-specific genes

Author
Travis, Adam,

Publication Date
1991

Peer reviewed|Thesis/dissertation

DEVELOPMENTAL REGULATION OF LYMPHOID-SPECIFIC GENES

by

ADAM TRAVIS

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOCHEMISTRY

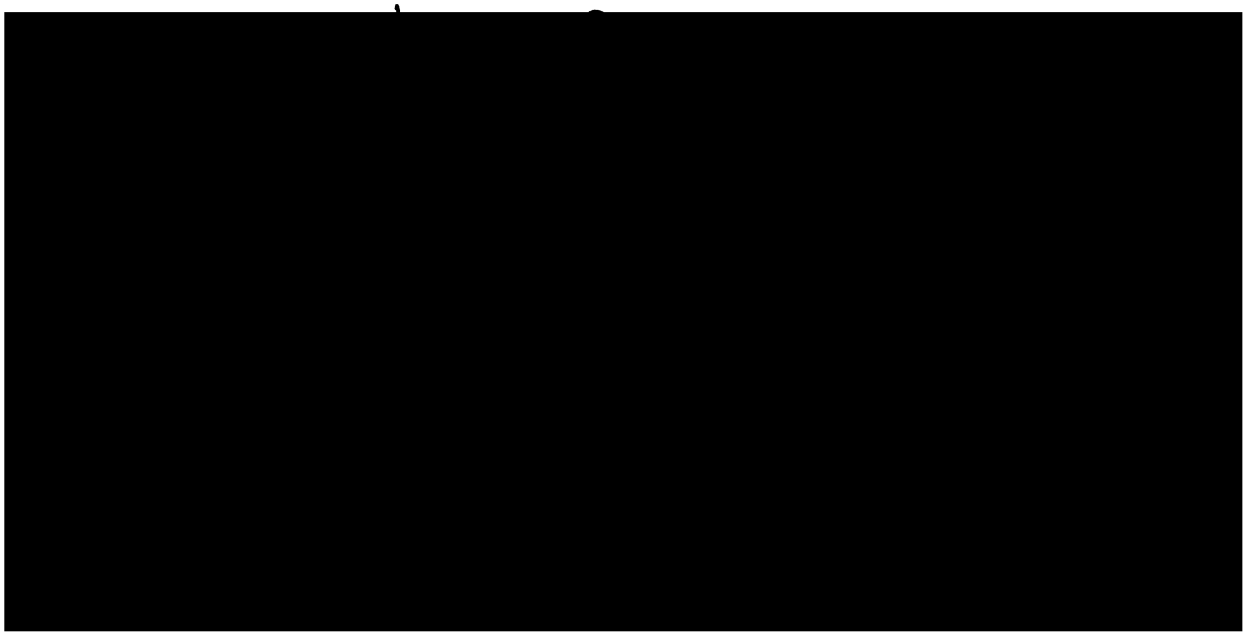
in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

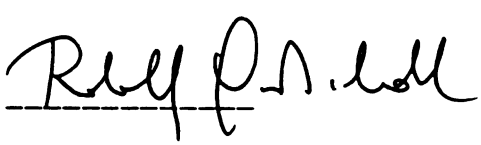
San Francisco



This thesis is dedicated to *Roy* and *Victoria*, my parents.

INVOLVEMENT OF COAUTHORS-ADVISOR'S STATEMENT

Adam Travis' thesis work resulted in the publication of four research articles. Moreover, two manuscripts to which Adam's work contributed significantly are being prepared for publication. All of these articles, including two that comprise chapters two and three of the thesis, involve coauthors. The authorship reflects the contributions of two and one other individual, respectively. For both chapters, however, the vast majority of the work was carried out by Adam. All of Adam's research, whether published with coauthors or described solely in the present thesis, was pursued independently. Adam's thesis work initiated two major research programs in the lab. I consider Adam's thesis work extremely successful and think that the involvement of collaborators in his work is the inevitable consequence of research in a rapidly progressing and competitive field.



Rudolf Grosschedl

ACKNOWLEDGMENTS

I am deeply grateful to many for helping me to reach this nodal point. First and foremost, I want to thank my advisor, Rudolf Grosschedl, for providing guidance and direction these past five years, and for helping me to mature, both as a scientist and a person. I also thank the members of the Grosschedl lab, both past and present, for making lab a fun and educational place to work. Of my collaborators, I am particularly grateful to James Hagman for his energy and commitment to understanding the workings of the *mb-1* promoter. I thank Adam Amsterdam and Carole Belanger for their tremendous job on the LEF-1 project, under pressure from our competitors. I am grateful to Keith Yamamoto and Richard Myers for critical reading of this thesis and for helpful discussions. Finally, I thank my friends and family, and in particular: Roland Bainton, Heather Colbert, Belen Garcia-Alvarado, Ken Goldman, Billy Forrester, Michael Lowenstein, Mary Marx, Alan Ruskin, Mary Lou Solbrig, Naomi and Roger Stoller, Evi Strauss, Sri Swangnam, Roy and Victoria Travis, and Gabriel and Robin Travis, for providing encouragement and inspiration when I needed it.

The text of chapters two and three of this thesis are reprints of papers as they appear in *Genes and Development* **5**: 880-894 and *Molecular and Cellular Biology* **11**: 5756-5766, respectively. The coauthors listed in these publications contributed figures 5, 6, and part of 7 in chapter two and figures 5, 6 and part of 7 in chapter 3.

DEVELOPMENTAL REGULATION OF LYMPHOID-SPECIFIC GENES

Adam Travis

ABSTRACT

Putative regulators of lymphoid-specific gene expression and differentiation were isolated by using two different approaches. First, several novel lymphoid-specific cDNA clones were isolated by differential library screening. Binding of their encoded polypeptides to DNA was then used as a criterion for identifying cDNA clones representing putative regulatory genes. Two lymphoid-specific cDNA clones were identified that encode polypeptides that bind to DNA-cellulose. One was further characterized and found to represent a gene, termed LEF-1 (lymphoid enhancer-binding factor-1), which is expressed specifically in pre-B and T cells. The LEF-1 protein contains a region of amino acid sequence homology to the chromosomal nonhistone protein HMG-1 and to several putative regulators of cell specialization, including the mammalian testis-determining factor SRY and fungal mating-type proteins. Immunocytochemical staining indicated that the LEF-1 protein is localized to the nucleus of pre-B cells. LEF-1 binds specifically to a sequence element in the T cell receptor (TCR) α enhancer. Maximal TCRα enhancer activity was found to parallel the cell type-specific expression pattern of LEF-1. Moreover, expression of recombinant LEF-1 in late stage B cells, which normally do not express it, increases TCRα enhancer function.

A second approach to the isolation of putative regulators of the lymphoid lineage began with a study of the cell type-specific regulation of the *mb-1* gene. The TATA-less *mb-1* promoter directs early B cell-specific transcription from multiple sites, paralleling endogenous mb-1 gene transcription. Analysis of the heterogeneity of transcription initiation led to the conclusion that upstream factor-binding sites in the mb-1 promoter define a region in which initiation of

transcription occurs at multiple sites. The binding sites for two ubiquitous nuclear factors positioned between -59 and -38 were found to be essential to *mb-1* promoter activity. A third functionally important binding site for an early B cell-specific nuclear factor is positioned between -180 and -160. This factor, termed EBF (early B cell factor), was purified from pre-B cells by sequence-specific DNA affinity chromatography. A 60 kilodalton polypeptide species copurified with EBF DNA-binding activity. Gel purification and renaturation of this species lead to the conclusion that it corresponds to EBF.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE:

CHAPTER TWO:

CHAPTER THREE:

CHAPTER FOUR:

INTRODUCTION

The developmental programs that direct a fertilized mammalian ovum to express the cell type diversity and spatial patterning of the adult organism must somehow be encoded in the four-letter alphabet of nucleic acids, the genetic material. A major challenge of developmental biology is to decipher these programs and to determine how they unfold from the information stored in the genetic material. By analogy with organisms that have been studied genetically (yeast, worm, fruit fly, etc.), mammalian development likely involves hierarchies of regulatory genes that are themselves regulated by spatial and temporal cues. The crucial site of regulation is likely to be at the level of transcriptional control of cell type-specific genes. Supporting this notion is the recent discovery of "master regulatory genes" such as *myoD*, which appear to specify particular cell types by encoding transcription factors for cell type-specific genes. Since these transcription factors are themselves cell type-specific, they can contribute to the cell type specificity of the sets of genes they regulate.

I. THE REGULATION OF TRANSCRIPTION INITIATION

Gene regulatory sequences

The notion that, in addition to the classical structural genes, there exist two other types of genetic determinants (regulator and operator) that control the rate of transfer of information from gene to protein by regulating the activity of the structural gene, was first proposed from genetic studies of enzyme induction and repression in *E. coli* (Jacob and Monod, 1961). A model was proposed in which the cytoplasmic product of the regulator gene (the repressor) could influence the rate of synthesis by the structural gene of a short-lived intermediate messenger

(mRNA) by interacting (or failing to interact) with an "operator" sequence located near the structural gene. Thirty years of experimentation in a wide variety of systems has substantiated and extended this model to include a host of prokaryotic and eukaryotic transcriptional regulators that function through binding regulatory DNA sequences.

The gene regulatory region from which transcription initiates, the promoter, often contains conserved sequence elements. In a large number of prokaryotic promoters, variations on the consensus sequences TATAAT and TTGACA are found around the -10 and -35 positions, respectively, where +1 is the transcription start site. These sequences are recognized by the RNA polymerase holoenzyme and allow the formation of a promoter-polymerase complex (von Hippel et al., 1982). In eukaryotic promoters, a TATA element resembling the -10 consensus sequence of prokaryotic promoters, is often present ~30 bp upstream of the transcription start site. TATA elements are recognized by the general transcription factor TFIID (Davison et al., 1983; Sawadogo and Roeder, 1985; Nakajima et al., 1988), and they have been shown to specify the site of transcription initiation (Grosschedl and Birnstiel, 1980; Grosveld et al., 1981; Breathnach and Chambon,1981) and to be critical for promoter activity in vitro (Wasylyk et al., 1980; McKnight and Kingsbury, 1982; Concino et al., 1984). However, a number of genes have been identified that have promoters lacking a TATA element. Some of these TATA-less promoters initiate transcription within a sequence element called an initiator, which appears to substitute for a TATA element (Smale and Baltimore, 1989; Smale et al., 1990). In addition to these basal elements, promoters usually contain other sequence elements that bind transcriptional regulators.

Additional regulatory regions termed enhancers stimulate transcription from linked promoters in a distance- and position-independent manner (Banerji et al., 1981; Moreau et al., 1981). Enhancers are usually associated with eukaryotic

genes and are often located several kilobases away from the promoter. They have been identified both upstream (Theisen et al., 1986) and downstream (Choi and Engel, 1986; Hesse et al., 1986) of the transcribed region, as well as within the transcribed region in intron sequences (Banerji et al., 1983; Gillies et al., 1983). Like promoters, enhancers contain sequence elements that bind transcription factors.

Prokaryotic transcriptional control

The analysis of over 100 *E. coli* promoters has revealed two families, one of which appears to have several features in common with eukaryotic promoters (reviewed in Gralla, 1991). The major family of *E. coli* promoters uses RNA polymerase holoenzyme containing sigma factor 70 (σ^{70}), has basal elements at -10 and -35, and recruits activators and repressors within or very close to the region of the basal elements. By contrast, a second, smaller family of *E. coli* promoters uses RNA polymerase containing σ^{54}, has basal elements around -12 and -24, and recruits activators at enhancer-like elements located at a distance from the basal elements. While the σ^{70} promoters require that an activator or repressor site be positioned to allow direct communication with polymerase binding the basal elements, σ^{54} promoters, like eukaryotic RNA polymerase II promoters, do not have this distance requirement. Activator elements of σ^{54} promoters can be moved kilobases away from the basal elements and retain functional activity (Reitzer and Magasanik, 1986; Buck et al., 1986; Birkmann and Bock, 1989).

Two key features of σ^{54} promoter function have implications for models of eukaryotic transcriptional control. First, a stable complex of core polymerase plus σ^{54} bound at the basal elements forms prior to activation of transcription (Sasse-Dwight and Gralla, 1988; Popham et al., 1989). Second, an activator bound at an upstream site contacts the basal element complex by looping out the intervening

DNA (Su et al., 1990). There is strong evidence that this contact between an activator and promoter-bound σ^{54} holoenzyme through loop formation catalyzes the transition to the open complex, which initiates transcription (Sasse-Dwight and Gralla, 1988; Popham et al. 1989; Wedel et al., 1990; Hoover et al., 1990). As discussed in the next section, there is evidence that both the formation of a stable pre-initiation complex of transcription factors bound close to the initiation site and the involvement of protein-protein contacts between activators bound at a distance and components of the pre-initiation complex are features of eukaryotic promoter function.

From an evolutionary perspective, regulation of transcription by distant enhancer-like elements has the advantage of greater flexibility, since these elements can be moved around without loss of function. In addition, since direct juxtaposition of regulatory elements and basal elements is not required, regulatory regions can be larger, allowing them to integrate multiple informational inputs. One potential disadvantage, however, is that regulation by distant elements appears to require that genes with different expression patterns be separated by long distances to avoid cross-regulation. This may account for the fact that σ^{54} promoters are rare in the compact genome of *E. coli* (Magasanik, 1989).

Eukaryotic transcriptional control

Both the transcription apparatus and the mechanisms of controlling it are considerably more complex in eukaryotes than in bacteria. There are three nuclear RNA polymerases, each composed of nine to fourteen subunits, several of which are common to two or all three enzymes (Lewis and Burgess, 1982). RNA polymerase I transcribes the large rRNA genes, RNA polymerase II transcribes the structural genes, and RNA polymerase III transcribes the 5S rRNA and tRNA genes and many small nuclear RNA genes.

Studies of the control of gene expression in eukaryotes have focused on the regulation of transcription initiation by RNA polymerase II. The regulatory sequences of many eukaryotic structural genes have been studied in detail, revealing certain common features. Eukaryotic enhancers and promoters often consist of multiple sequence elements that have specific regulatory functions and bind specific nuclear factors. Elements have been identified that contribute to the cell type specificity, inducibility, or ubiquitous activation of the gene they regulate. In some cases, the function of an individual element is strongly influenced by the adjacent sequences present in the regulatory region. This context dependence suggests that the multiple sequence elements of an enhancer or promoter can interact with each other, which would greatly increase the number of regulatory possibilities generated by shuffling elements (Voss et al., 1986). In certain cases, context dependence has been shown to arise from interactions of the transcription factors which bind adjacent elements. For example, adjacent elements that function synergistically to activate transcription may do so by the cooperative binding of transcription factors (Poellinger et al., 1989). The interaction of factors can also result in negative synergism between adjacent elements. In certain contexts, the binding of a particular factor to an element has either negative or positive effects on transcription, depending on which factor binds an adjacent element (Diamond et al., 1990).

What is the mechanism by which eukaryotic enhancers increase the rate of transcription initiation? One model involves protein-protein contact between factors bound at enhancers and components of the basic transcription machinery bound at the promoter (Ptashne, 1986). This would require looping of the intervening DNA between the enhancer and the promoter. In a second model, RNA polymerase II or a transcription factor binds the enhancer and then scans along the DNA until it reaches the promoter. Consistent with the looping model, an enhancer

attached to a promoter via a protein bridge was found to be capable of stimulating transcription *in vitro* (Muller et al., 1989).

The general factors involved in RNA polymerase II transcription

A number of general transcription factors have been found to be necessary for accurate RNA polymerase II transcription initiation (reviewed in Saltzman and Weinmann, 1989; Lillie and Green, 1989). These include RNA polymerase II itself and other chromatographically defined factors including TFIIB, TFIID, and TFIIE/F. Another factor, termed TFIIA, has been implicated as a general transcription factor, but the requirement for it is not strict. The general factors appear to assemble on the promoter in a defined order to form a preinitiation complex (Buratowski et al., 1989). A complex of TFIID and the promoter initiates the assembly by recruiting TFIIB, which leads to the recruitment of RNA polymerase II and TFIIE/F. While the general factors alone are sufficient for preinitiation complex assembly and accurate transcription initiation *in vitro*, transcriptional activity is greatly stimulated by activators that bind specific promoter sequences.

The question arises whether promoter-specific transcriptional activators function by facilitating the assembly of the preinitiation complex, or by increasing the activity of the complex after it has assembled. For one type of transcriptional activator containing an acidic domain, the mechanism of action appears to involve an increase in the rate of preinitiation complex assembly (Lin and Green, 1991). This suggests that in the absence of an activator, preinitiation complex assembly stalls at a rate-limiting step. The recruitment of TFIIB, which may act as a "bridging" molecule between the TFIID-TATA element complex and RNA polymerase II (Buratowski et al., 1989), appears to be such a rate-limiting step. The acidic transcriptional activator appears to increase the rate of preinitiation complex assembly by accelerating this step (Lin and Green, 1991). Is TFIIB the direct

target of acidic activators? Specific interactions between an acidic activation domain and both TFIIB and TFIID have been observed, but the interaction with TFIIB appears to be stronger (Lin and Green, 1991; Stringer et al., 1990). It is conceivable that transcriptional activators have more than one target.

TFIID contains a polypeptide that binds to TATA elements, termed TATA-binding protein (TBP), the gene for which has been cloned from yeast, Drosophila, and human (reviewed in Greenblatt, 1991). Studies using recombinant TBP in place of TFIID indicate that TFIID fractions contain other components, termed coactivators, required for the function of transcriptional activators (Pugh and Tjian, 1990). Purified TFIID is much larger than TBP and appears to be a multisubunit protein. TBP in Drosophila TFIID is tightly associated with at least six polypeptides termed TBP-associated factors (TAFs) (Dynlacht et al., 1991). Purified TAFs added to TBP can mediate transcriptional activator function. Thus one or more of the TAF polypeptides can function as coactivators.

Under certain conditions, the first step in preinitiation complex assembly, the binding of TFIID to the promoter, appears to be facilitated by TFIIA (reviewed in Sumimoto et al., 1990). This effect may reflect the cooperative binding of TBP and TFIIA to the TATA element (Meisterernst and Roeder, 1991). A family of TFIIA-related factors has been identified that bind cooperatively with TBP to TATA elements. One of these TFIIA-related factors may act as a negative regulator of the subsequent steps in preinitiation complex assembly (Meisterernst and Roeder, 1991).

Transcription factors that bind to specific regulatory sequences

In addition to the general factors required for basal RNA polymerase II transcription, a host of factors that bind to functionally important sequence elements in promoters and enhancers have been identified. Taking advantage of their

sequence-specific DNA binding properties, techniques for purifying these factors (Kadonaga and Tjian, 1986) and for screening cDNA expression libraries (Singh et al., 1988; Vinson et al.,1988) have allowed many genes encoding mammalian transcription factors to be cloned and characterized.

Mutational analyses of these cloned transcription factors have identified structural domains that mediate sequence-specific DNA binding and transcriptional activation of target promoter constructs (Mitchell and Tjian, 1989). Amino acid sequence comparisons between these factors and other proteins have revealed that many of the structural domains involved in DNA binding can be grouped into evolutionarily conserved families. Some of these families include genes identified genetically that participate in processes as diverse as cell-type determination, pattern formation, and the regulation of cell growth. Many of these genetically identified genes are now known to encode sequence-specific DNA binding proteins that regulate transcription. A summary of several of the known families of sequence-specific DNA binding proteins is given in the next section.

Although the activation domains of transcription factors are more difficult to predict than DNA binding domains based on amino acid sequence alone, several families of domains rich in particular amino acids have been defined. Studies of the yeast transcription factors GAL4 and GCN4 (Ptashne, 1988; Hope and Struhl, 1986) revealed a family of activation domains rich in acidic amino acids and believed to be capable of forming amphipathic α-helical structures. Acidic activation domains have also been identified in a few mammalian transcription factors (Mitchell and Tjian, 1989). Deletion analysis of the transcription factor Sp1 revealed two activation domains that contain ~25% glutamine and very few charged amino acids. Similar glutamine-rich activation domains have been identified in several yeast, Drosophila, and mammalian transcription factors (Mitchell and Tjian, 1989). A third type of activation domain rich in prolines was

identified in the transcription factor CTF/NF-1. Proline-rich regions are also present in several other mammalian transcription factors (Mitchell and Tjian, 1989). It is believed that all three of these activation domain families function by contacting other proteins.

Families of sequence-specific DNA binding proteins

Six different families of DNA binding domains found in mammalian transcription factors (Zn-fingers, homeodomain, basic leucine zipper, basic helix-loop-helix, basic helix-loop-helix leucine zipper, and HMG domain) are discussed in this section. For most, if not all of these domain types, examples can be found among the regulatory proteins of a wide range of eukaryotic taxa, from yeast to man. This high degree of conservation suggests that these structural solutions to the problem of generating a protein capable of forming a high affinity sequence-specific complex with DNA, were attained early in evolution, and have been extensively exploited ever since. In addition to their direct role in contacting DNA, several of these domains are involved in the formation of homodimers and/or heterodimers with other members of the same domain family. For many of the families, it is this dimer formation that creates the DNA-binding structure.

The term "zinc fingers" was originally used to describe the DNA binding structures in the RNA polymerase III transcription factor TFIIIA, which binds to regulatory sequences of the 5S RNA gene (Miller et al., 1985). Each finger consists of ~30 amino acids and contains two cysteines and two histidines that nucleate the structure of the domain by tetrahedrally coordinating a Zn^{2+} ion. Zn-fingers are present in the mammalian transcription factor Sp1 (Kadonaga et al., 1987) and many other eukaryotic regulatory proteins. Genes encoding proteins with as many as 37 putative Zn-fingers have been identified (Altaba et al., 1987). A second type of Zn-coordinating structure is found in the DNA binding domains of the steroid

hormone receptors. These domains consists of a pattern of eight cysteines which coordinate two Zn^{2+} ions, each with tetrahedral geometry (Freedman et al., 1988). Unlike the Zn-fingers found in TFIIIA, which form independent, stable structures, each of which contribute to DNA binding (Pavletich and Pabo, 1991), the Zn-fingers of the steroid hormone receptors appear to fold together to form a unified globular domain. The crystal structure of a glucocorticoid receptor Zn-fingers-DNA complex revealed a dimeric globular structure with an α-helical surface of each subunit that makes specific base contacts in the major groove of DNA, and a surface formed from part of the second finger that mediates dimerization (Luisi et al., 1991). A third type of Zn-coordiating DNA binding domain is exemplified by the yeast regulator GAL4 (Vallee et al., 1991).

The homeodomain is a DNA binding structure originally identified in a set of Drosophila proteins controlling segmentation in the developing embryo (McGinnis et al., 1984; Poole et al., 1985). Homeodomains have highly conserved amino acid sequences of ~60 residues and are found in many proteins that are implicated in mammalian developmental control (Kessel and Gruss, 1990) and that regulate transcription (Scott et al., 1989). Although homeodomains can be grouped into subfamilies based on amino acid sequence comparison, it is likely that they all contain a helix-turn-helix structure similar to that found in several prokaryotic repressors (Pabo and Sauer, 1984). The crystal structure of an engrailed homeodomain-DNA complex revealed three α helices that pack together so that the exposed, hydrophilic face of helix 3 (the "recognition helix") interacts with basepairs in the major groove of DNA (Kissinger et al., 1990). A number of homeodomain proteins, including the mammalian transcription factors Oct-1, Oct-2, and Pit-1 and the *C. elegans* developmental regulator Unc-86 comprise a subfamily referred to as POU-factors (for Pit-1, Oct, and Unc-86) (Herr et al., 1988). The homeodomain of these factors is part of a larger conserved, bipartite

domain of ~150 amino acids, termed the POU domain. The POU domain consists of a conserved POU-specific domain (POU_S) of ~70 amino acids, a poorly conserved linker region of ~ 20 amino acids, and a conserved POU-specific homeodomain (POU_{HD}) of ~60 amino acids. Both POU_S and POU_{HD} are required for high-affinity, site-specific DNA binding and, for a subset of POU factors on a subset of binding sites, the POU domain mediates dimerization (Rosenfeld, 1991).

The basic leucine zipper domain (bZIP) was first identified as a 60 amino acid region of the transcription factor C/EBP, which was found to share homology with a number of other proteins, including the products of the *jun* and *fos* oncogenes (Landschulz et al., 1988). The bZIP domain is a bipartite structure that consists of a basic region of ~30 amino acids that contacts DNA, immediately followed by a dimer-forming region containing four leucines positioned at intervals of seven amino acids. Considerable evidence indicates that the leucines are situated on an α helix to form the hydrophobic dimerization interface termed the leucine zipper (O'Shea et al., 1989; Abel and Maniatis, 1989; Jones, 1990). The leucine zipper helices of the dimer appear to intertwine around one another in a parallel orientation (O'Shea et al., 1989). One model for the interaction of bZIP proteins with DNA suggests that the two basic regions of a dimer form α helices that track in opposite directions along the major groove to form a "scissors grip" around DNA (Vinson et al., 1989). When the leucine zipper of the bZIP domain of the yeast regulatory protein GCN4 was replaced by a disulfide bond, the disulfide-linked dimer could bind DNA sequence-specifically (Talanian et al., 1990). This indicates that the basic region is sufficient for sequence-specific DNA binding and suggests that the leucine zipper functions mainly to mediate protein dimerization. In addition to forming homodimers, members of the bZIP protein family are known to heterodimerize. Examples include the products of the *jun* and *fos* cellular oncogenes, which dimerize to form the transcription factor AP-1 (Curran and

Franza, 1988), and members of the ATF/CREB family of cAMP responsive factors (Ziff, 1990).

The basic helix-loop-helix domain (bHLH) was first identified as a DNA binding and dimer forming structure in the E12 and E47 immunoglobulin enhancer binding proteins, the products of the *daughterless* and *achaete-scute* genes of *Drosophila*, and the MyoD and Myc proteins (Murre et al., 1989a). The bHLH domain consists of a basic region of ~15 amino acids just amino terminal to a region of 35-50 amino acids that is believed to form two amphipathic α-helices with an intervening loop. Although the structure of the bHLH domain is not known, mutagenesis has revealed that the basic region is required for DNA binding and the HLH region is required for dimerization (Davis et al., 1990; Voronova and Baltimore, 1990). Several bHLH proteins expressed in all tissues (E12, E47, and daughterless) can form heterodimers with bHLH proteins expressed in specific tissues (MyoD and achaete-scute), and often bind, both as homodimers and heterodimers, to the same core consensus sequence CA--TG (Murre et al., 1989b; Jones, 1990; Blackwell and Weintraub, 1990). The formation of hetero-oligomers of MyoD with E12/E47-like proteins occurs in vivo and may be essential to the muscle-specific regulatory role of MyoD (Lassar et al., 1991). Interestingly, HLH proteins that lack a basic region have been identified (Jones, 1990). These proteins appear to act as negative regulators by interacting with bHLH proteins to form heterodimers that fail to bind DNA (Benezra et al., 1990).

The basic helix-loop-helix leucine zipper domain (bHLH-ZIP) was originally defined by the Myc family of proteins, and subsequently was identified in the immunoglobulin enhancer binding proteins TFE3 and TFEB, as well as the transcription factors USF and AP4 (Luscher and Eisenman, 1990) This domain consists of a bHLH region, just amino terminal to a heptad array of 4-5 leucines. For USF and AP4, deletion mutants have revealed that both the HLH and zipper

regions are required for dimer formation and DNA binding (Gregor et al., 1990; Hu et al., 1990). For c-Myc, a DNA-binding site for the bHLH-ZIP domain was identified by reiterative selection and amplification of partially randomized oligonucleotides containing the bHLH binding site consensus CA--TG (Blackwell et al., 1990). The binding site sequence identified, CACGTG, is likely recognized by homodimers. Relatively high concentrations of the c-Myc protein fragment were required, which may reflect the fact that c-Myc homodimers are not readily formed. This suggests that c-Myc must have a partner that heterodimerizes with it and allows it to bind DNA with higher affinity. By screening a cDNA expression library with the bHLH-ZIP domain of the c-Myc protein, a novel bHLH-ZIP protein, termed Max, was identified (Blackwood and Eisenman, 1991). Max was also identified by a degenerate polymerase chain reaction using primers derived from the cMyc bHLH-ZIP amino acid sequence, and termed Myn (Prendergast et al., 1991). Max/Myn specifically associated with c-Myc, N-Myc, and L-Myc proteins, but not with bHLH, bZIP, or other bHLH-ZIP proteins. The cMyc-Max/Myn complex bound specifically to the c-Myc binding site with significantly higher affinity than either c-Myc or Max/Myn alone. The DNA binding activity of the complex required the basic region and HLH-ZIP dimerization domain of c-Myc (Blackwood and Eisenman, 1991). Coexpression of Max/Myn with c-Myc augmented focus formation in a transformation assay, suggesting that Max/Myn may facilitate the function of c-Myc in vivo (Prendergast et al., 1991).

The HMG domain is a DNA binding structure of ~80 amino acids first identified as a region of homology between the human RNA polymerase I transcription factor hUBF and the high mobility group protein HMG1 (Jantzen et al., 1990). While hUBF is a sequence-specific DNA binding protein, HMG1 binds DNA relatively nonspecifically and is not known to regulate specific genes. hUBF has four HMG domains, deletion of which prevents DNA binding. A variety of known or candidate

regulatory proteins containing a single HMG domain have been identified. Examples are the mating-type proteins mat-Mc of *S. pombe* (Kelly et al., 1988) and mt a-1 of *N. crassa* (Staben and Yanofsky, 1990), the putative mammalian testis-determining factor SRY (Sinclair et al., 1990; Gubbay et al., 1990) and the pre-B and T lymphocyte-specific enhancer-binding factor LEF-1 (Travis et al., 1991; Waterman et al., 1991). For LEF-1, the HMG homology was shown to be the DNA binding domain (Giese et al., 1991). Four autosomal genes encoding proteins with HMG domains very similar to that of SRY are expressed in the early mouse embryo and appear to be part of a larger SRY-related gene family (Gubbay et al., 1990). Similarly, the T lymphocyte-specific transcription factor TCF-1 (van de Wetering et al., 1991) has an HMG domain 98% identical to that of LEF-1/TCF-1α, and recognizes a very similar and possibly overlapping set of binding sites. These results indicate that subfamilies of HMG domain proteins exist. The structure of the HMG domain and how it interacts with DNA are unknown.

Additional DNA binding domain types include the ets domain family, exemplified by the mammalian transcription factors Ets-1, Ets-2 and PU.1 (Watson et al., 1988; Gunther et al., 1990; Klemsz et al., 1990), the highly basic DNA binding domain of the mammalian transcription factor CTF-1/NF-1 (Mitchell and Tjian, 1989), the DNA binding and putative dimerization structure of c-Rel and the subunits of NF-κB (Ghosh et al., 1990; Nolan et al., 1991), and the DNA binding domains of the liver-enriched transcription factors HNF-3α, β, and γ, which share homology with a domain of the *Drosophila* homeotic protein fork head (Lai et al., 1990; Lai et al., 1991).

II. MAMMALIAN CELL TYPE-SPECIFIC GENE EXPRESSION

Mammal cell type-specific DNA binding proteins

A summary of mammalian sequence-specific DNA binding factors which have a cell type-specific expression pattern is given in Table 1. The factors are grouped on the basis of exclusive or preferential expression in hematopoietic, liver, muscle, brain, thyroid, or pancreatic endocrine tissue. Most of the factors shown in Table 1 were identified as cell type-specific nuclear factors that bind DNA sequence elements in the enhancers and/or promoters of cell type-specific genes. For a few, such as the muscle cell lineage determining factor MyoD, cell type-specific cDNA clones were first isolated and subsequently shown to encode regulators of cell type-specific target genes. Interestingly, the cell type distributions of several of the factors shown in Table 1 do not exactly parallel those of their suspected target genes. In general, the factors appear to have a broader expression pattern than their target genes, supporting the notion that they are necessary but not sufficient for target gene expression. Members of the MyoD family (MyoD, Myogenin, Myf-5, and MRF4/Myf-6/Herculin) appear to be exceptions.

MyoD: A lineage determining transcription factor

The discovery of a single gene capable of converting a variety of differentiated cell types into muscle (Lassar et al., 1986; Davis et al., 1987; Weintraub et al., 1989; Lin et al., 1989) has lead to many insights and questions about the mechanisms that specify diverse cell types. The myoD gene is stably expressed only in skeletal muscle and its precursors and encodes a protein that activates transcription of muscle-specific genes by directly binding to their control regions (Weintraub et al., 1991a; Weintraub et al., 1991b). The MyoD protein contains a bHLH domain required for dimerization and DNA binding (discussed above) and

Table 1. Cloned mammalian cell type-specific DNA binding proteins

Factor	Cell type distribution	Family	Suspected target genes	Binding site sequences	References
	HEMATOPOIETIC LINEAGES				
Oct 2 (OTF-2)	B cells, some T cells	pou-homeo	immunoglobulin genes B29, MHC class II genes	ATTTGCAT	Muller et al., 1988; Scheidereit et al., 1988; Staudt et al., 1988
PU.1	B cells, macrophages	ets	MHC class II gene I-Aβ	GAGGAA	Klemsz et al., 1990
LEF-1 (TCF-1α)	pre-B and T cells	HMG	TCRα, CD4, CD3-δ, CD3-γ	$^{C}/_{T}$CTTTGAA	Travis et al., 1991; Waterman et al., 1991
TCF-1	T cells	HMG	CD3-ε, TCRα	GCTTTGTT	van de Wetering et al., 1991
GATA-3	T cells	Zn-finger	TCRα, TCRδ	AGATAG	Ho et al., 1991; Joulin et al., 1991
GATA-1 (GF-1, Eryf1, NF-E1)	erythroid cells, megakaryocytes	Zn-finger	α-, β-, and γ-globin, porphobilinogen deaminase	$^{A}/_{T}$GATA$^{A}/_{G}$	Tsai et al., 1989; Evans and Felsenfeld, 1989; Yamamoto et al., 1990
	GONAD				
SRY	testis	HMG	unknown	CCATTGTT	Sinclair et al., 1990; Gubbay et al., 1990
	LIVER-ENRICHED				
LF-B1 (HNF-1α)	liver, kidney > intestine > spleen	homeo	albumin, transthyretin α1-antitrypsin, α-fetoprotein	GTTAATNATTAAC	Frain et al., 1989; Baumhueter et al., 1990
LAP	nuclear protein: liver RNA: lung > liver, spleen	bZIP	albumin	(TGATTTTGTAATGGG)	Descombes et al., 1990
DBP	nuclear protein: liver RNA: liver > spleen, brain	(basic domain, no ZIP)	•	•	Mueller et al., 1990
HNF-3α	liver, lung	(related to *Drosophila* homeotic gene *fork head*)	transthyretin, α1-antitrypsin	TATTGA$^{C}/_{T}$TT$^{A}/_{T}$G	Lai et al., 1990
HNF-3β	• •	• •	•	•	Lai et al., 1991
HNF-3γ	liver >> intestine	• •	•	•	•
HNF-4	liver, kidney, intestine	Zn-finger (hormone receptor family)	transthyretin, apolipoprotein CIII	$^{G}/_{T}$GC$^{A}/_{T}$A$^{A}/_{G}$G$^{G}/_{T}$T$^{T}/_{C}$CA$^{T}/_{C}$	Sladek et al., 1990
	MUSCLE				
MyoD	skeletal muscle	bHLH	muscle α-actin, myosin, troponin-T muscle creatine kinase, MyoD family	CANNTG	Davis et al., 1987
Myogenin	• •	•	• • • •	•	Edmondson and Olson, 1989, Wright et al., 1989
Myf-5	• •	•	• • • •	•	Braun et al., 1989
MRF4 (Myf-6, Herculin)	• •	•	• • • •	•	Rhodes and Konieczny, 1989; Braun et al., 1990
	BRAIN				
GHF-1 (Pit-1)	pituitary	pou-homeo	growth hormone, prolactin	AATATNCAT	Bodner et al., 1988; Ingraham et al., 1988
	THYROID-ENRICHED				
TTF-1	thyroid, lung	homeo	thyroglobulin, thyroperoxidase	GNNCACTCAAG	Guazzi et al., 1990
	PANCREATIC ENDOCRINE				
Isl-1	islets of pancreas	Cys-His/homeo	insulin 1	TTAATAATCTAA	Karlsson et al., 1990

appears to require heterodimerization with a ubiquitously expressed protein related or identical to E12/E47 for activation of the myogenic program (Lassar et al., 1991).

Heterodimers of MyoD with the E12 product of the E2A gene bind to the muscle-specific creatine phosphokinase (MCK) enhancer with >10-fold higher affinity than MyoD homodimers (Murre et al., 1989b). Although homodimers and heterodimers recognize the same core consensus sequence (CA--TG), they prefer different nucleotides at other binding site positions (Blackwell and Weintraub, 1990). Several lines of evidence indicate that stable MyoD expression and subsequent heterodimer formation represents a nodal point in myogenesis (Weintraub et al., 1991b; Lassar et al., 1991). How is arrival at this nodal point regulated during development?

Regulating the regulator

Although the mechanism of MyoD induction in muscle cell precursors is unknown, much is known about mechanisms that likely serve to stabilize the commitment to MyoD expression and the muscle cell lineage. In *Xenopus laevis*, the myoD gene is initially transcribed at low levels throughout the early embryo prior to mesoderm induction (Rupp and Weintraub, 1991). Subsequently, myoD expression is stabilized only in a specific subset of mesodermal cells commited to the skeletal muscle cell lineage (Hopwood et al., 1989). Two control mechanisms, the down-regulation of an inhibitor and positive auto-activation, likely contribute to the irreversibility of this commitment to myogenesis.

Proliferating myoblasts in tissue culture express MyoD protein, yet do not differentiate as long as serum is present (Davis et al., 1987). In vivo footprinting of the MCK enhancer reveals that the MyoD binding site is occupied in differentiated myotubes but not in myoblasts (Mueller and Wold, 1989). Thus the activity of the

MyoD protein is inhibited prior to differentiation. This negative regulation appears to involve the product of the Id (inhibitor of differentiation) gene (Benezra et al., 1990). The Id protein contains an HLH domain that allows it to form heterodimers with MyoD and E2A proteins, but is lacking a basic domain, making dimers that contain Id unable to bind DNA. Id is expressed in most cell types, including myoblasts where it is down-regulated 10- to 20-fold upon differentiation. Id inhibits MyoD-mediated activation of a muscle-specific reporter gene and retards myogenic differentiation when expressed from a viral LTR (Weintraub et al., 1991a). Since Id binds with higher affinity to E2A proteins than to MyoD, it may inhibit differentiation by sequestering E2A proteins, preventing heterodimer formation and the activation of muscle-specific genes.

The myoD gene is a member of a multigene family (myoD, myogenin, myf-5, mrf4/herculin/myf-6) each member of which can activate myogenesis (reviewed in Weintraub et al., 1991a). Transfection of any one of these genes leads to the activation of most or all of the others, as well as the endogenous copy of itself. This cross- and auto-activation insures that once myogenesis is induced, large amounts of myogenic regulatory proteins are produced. This may serves to reinforce the commitment to the muscle cell lineage.

MyoD: A paradigm?

Do regulatory equivalents of myoD exist for other cell lineages? The *achaete-scute* gene complex of *Drosophila* seems to encode a master regulator of neurogenesis that shares several properties with MyoD (reviewed in Weintraub et al., 1991a). Mutations in achaete-scute prevent the differentiation of peripheral neurons (Romani et al., 1989) and additional genetic evidence suggests an interaction with a second gene, *daughterless* (*da*) (Caudy et al., 1988), which is involved in sex determination and myogenesis, as well as neurogenesis (Cline, 1989; Parkhurst et

al., 1990). Both *achaete-scute* and *da* encode bHLH proteins and *da* is homologous to E2A and expressed in many cell types. A third Drosophila gene, *extra macrochaetae* (*emc*) appears to be an analog of Id and genetic evidence suggests that it is a negative regulator of *achaete-scute* (Ghysen and Dambly-Chaudiere, 1989). Like Id, emc encodes an HLH protein that lacks a basic region and is expressed in most cells (Ellis et al., 1990; Garrell and Modolell, 1990; Parkhurst et al., 1990). Thus, at least in *Drosophila*, sets of interacting HLH proteins are involved both in neurogenesis and myogenesis: MyoD and *achaete-scute* specify cell types, Id and *emc* provide negative regulatory functions, and E2A and *da* appear to have effector functions.

The signal to specify a particular cell type need not always be propagated through the activation of a single regulatory entity such as the myoD gene family. An alternative mechanism could involve combinations of less cell type-specific regulatory genes that act together to specify a given cell type (Weintraub et al., 1991a).

III. LYMPHOCYTE DIFFERENTIATION

The B and T lymphocyte lineages

Lymphocyte differentiation converts progenitor cells into highly specialized cells of either the B or the T cell lineage. Both cell lineages involve multiple stages of differentiation defined by changes in the expression patterns of their antigen receptor genes. The B cell lineage can be divided into three major stages. Pre-B cells represent the earliest stage in which the assembly of a functional immunoglobulin (Ig) μ heavy-chain gene from variable and constant gene segments has occurred and μ heavy-chain protein is expressed. Subsequently, the Ig κ or λ light-chain gene is functionally assembled and expressed in B cells, which display

antibody on their surface. Together, pre-B and B cells represent the early, antigen-independent stages of B cell differentiation. After B cells encounter specific antigen, they terminally differentiate into antibody-secreting plasma cells (Blackwell and Alt, 1988; Cooper and Burrows, 1989). In a similar manner, T cells differentiate in stages, beginning with the sequential rearrangement of their T cell antigen receptor (TCR) β- and α-chain genes. Early T cells express the TCR and both the CD4 and CD8 cell surface molecules. These cells differentiate into helper- or cytotoxic-T cells expressing the TCR and either CD4 or CD8, respectively (Davis and Bjorkman, 1988; Blackman et al., 1990).

The availability of murine cell lines representing the various stages of B cell differentiation, as well as several murine and human T cell lines, makes the lymphocyte lineage a tractable system for the study of cell specialization. Most studies have aimed at identifying the *cis*-regulatory sequences and *trans*-acting factors controlling the genes that define the differentiated B and T cell phenotypes. Specific expression of Ig genes in B cells and TCR genes in T cells was shown to depend on transcriptional enhancers located within Ig genes in introns and 3' to the TCR genes (Gillies et al., 1983; Banerji et al., 1983; McDougall et. al, 1988; Ho et al., 1989; Winoto and Baltimore, 1989). Nuclear factors were identified that recognize sequence elements within the Ig and TCR enhancers and the genes encoding some of these enhancer-binding factors have been cloned and characterized. To date, four genes (Oct-2, LEF-1, TCF-1, and GATA-3) have been isolated which encode Ig or TCR enhancer binding factors that are strictly lymphoid-specific. The Oct-2 (OTF-2) gene is expressed in B cells and a subset of T cells and encodes a factor that binds the OCTA site found in the IgH enhancer and in most Ig promoters (Muller et al., 1988; Scheidereit et al., 1988; Staudt et al., 1988). The LEF-1 (TCF-1α) gene is pre-B and T cell-specific and encodes a factor that binds to the TCR α enhancer (Travis et al., 1991; Waterman et al., 1991). The TCF-1 gene is T cell-

specific and encodes a factor with a DNA binding domain that is 98% identical to that of LEF-1 and can recognize the same binding site in the TCR α enhancer and in the regulatory regions of other T cell-specific genes (van de Wetering et al., 1991; Oosterwegel et al., 1991). The GATA-3 gene is T cell-specific and encodes a factor that binds elements in both the TCR α and TCR δ enhancers (Ho et al., 1991; Joulin et al., 1991). Additional information on these factors is shown in Table 1. Because of their lymphoid-specific expression patterns, these four genes are good candidates for important regulatory participants in the lymphocyte lineage.

Temporal regulation during B cell differentiation

In addition to Ig and TCR genes, B and T cell function requires the expression of a variety of other lymphoid-specific genes. Many of these genes are temporally regulated during B and/or T cell differentiation. Several examples of genes that are temporally regulated during B cell differentiation are given in this section.

In both pre-B and pre-T cells, assembly of Ig and TCR genes is mediated by a shared recombination activity. This activity is likely encoded by the pre-B and pre-T cell-specific recombination activating genes RAG1 and RAG2 (Schatz et al., 1989; Oettinger et al., 1990). Another pre-B and pre-T cell-specific gene involved in Ig and TCR gene assembly encodes terminal deoxynucleotidyl transferase (TdT) (Landau et al., 1984). TdT is a template-independent DNA polymerase that inserts nucleotides at the junctions between joining gene segments during Ig and TCR gene rearrangements (Landau et al., 1987). These insertions function to increase diversity of the Ig and TCR repertoires.

In B cells, the expression of Ig on the cell surface requires the Ig-associated chains IgM-α and Ig-β (Venkitaraman et al., 1991), which are encoded by the early B cell-specific *mb-1* gene and B cell-specific B29 gene, respectively (Sakaguchi et al., 1988; Hermanson et al., 1988). The mb-1 gene shuts off during the transition

from B cells to plasma cells, consistent with the loss of surface expression of Ig. Finally, the pre-B cell-specific λ_5 and V_{preB} genes (Sakaguchi et al., 1986; Kudo and Melchers, 1987) appear to encode the Ig ω and ι surrogate light-chain components and may have a pre-B cell-specific signalling role as part of a cell surface complex with Ig μ heavy-chain (Cherayil and Pillai, 1991).

The promoters of several of these genes have been characterized (Hermanson et al, 1989; Smale and Baltimore, 1989; Travis et al., 1991b; Hagman et al., 1991; Lo et al., 1991), but the mechanisms underlying their cell type- and stage-specific expression are not yet known.

CHAPTER ONE:

The isolation and characterization of novel lymphoid-specific cDNA clones.

ABSTRACT

With the aim of identifying regulatory participants in lymphocyte gene expression, fourteen independent lymphoid-specific cDNA clones were isolated by differential library screening and seven were found to be novel. The genes represented by these cDNA clones were found to have distinct cell type- and stage-specific RNA expression patterns within the B and T cell lineages. Polypeptides derived from the seven novel cDNAs were generated *in vitro* and two, termed GL1 and FB1, were found to bind to DNA-cellulose. The putative lymphoid-specific regulatory role of the genes represented by the GL1 and FB1 cDNAs is considered.

In addition, the sequence of a putative promoter of the gene represented by the GL1 cDNA (termed LEF-1) was determined. Putative transcription initiation sites for the LEF-1 gene were determined by primer extension. Finally, the structure of the major open reading frame of another lymphoid-specific cDNA, termed DL4, was determined and appears to encode a novel transmembrane protein.

INTRODUCTION

Most studies aimed at identifying regulators of cell type-specific gene expression in mammals have involved the identification of cell type-specific nuclear factors that bind to sequence elements in the enhancers and/or promoters of cell-type specific genes. This approach has lead to the cloning and characterization of several cell type-specific regulatory genes, either through the purification of such nuclear factors (Kadonoga and Tjian, 1986) or the screening of cDNA expression libraries with the binding sites for such nuclear factors (Singh et al., 1988; Vinson et al., 1988). Not all cell type-specific promoters and enhancers pocess obvious binding sites for cell type-specific nuclear factors. In fact, most nuclear factors identified as binding cell type-specific regulatory sequences have a ubiquitous expression pattern. In some cases, the binding site for a cell type-specific nuclear factor also binds a factor expressed in other cell types. For example, the binding site for the lymphoid-specific factor Oct-2 (Muller et al., 1988; Scheidereit et al., 1988; Staudt et al., 1988) found in the regulatory sequences of immunoglobulin genes, is also bound by the ubiquitous factor Oct-1 (Singh et al., 1986; Scholer et al., 1989), and the binding site for the liver-enriched factor HNF-1α (Frain et al., 1989; Baumhueter et al., 1990), found in the regulatory regions of several liver-specific genes, is also bound by the non-liver-specific factor HNF-1β (Mendel et al., 1991). This represents a potential obstacle to the identification of cell type-specific DNA binding factors since the binding of a ubiquitous factor to the same site could obscure the detection of the binding of a cell type-specific factor.

We have used and alternative strategy for the cloning of genes encoding putative lymphocyte-specific regulatory factors. Differential library screening was used to identify several novel lymphocyte-specific cDNA clones. Binding of the encoded polypeptide to DNA was then used as a criterion to identify cDNA clones that encode

putative regulatory proteins. Lymphocyte specificity is thus ensured prior to screening for the ability to bind DNA. Two cDNA clones, termed GL1 and FB1, were found to encode DNA binding proteins. Further characterization of the GL1 clone is given in Chapter Two.

MATERIALS AND METHODS

Isolation of cDNA clones

Putative lymphoid-specific cDNA clones were isolated from a murine 70Z/3 pre-B cell library in λgt11 (Ben-Neriah et al., 1986). The clones were identified by screening duplicate bacteriophage lifts with [α^{32}P]dCTP-labelled first strand cDNA generated from either 70Z/3 or MEL (murine erythroleukemia) poly(A)$^+$ RNA. Of ~200,000 recombiant (lacZ$^-$) bacteriophage plaques, 10,000 were picked as 16 pools of ~600 which hybridized with the labelled 70Z/3 cDNA but failed to hybridize with the labelled MEL cDNA. These pools were rescreened by the same method using labelled cDNA derived from the murine T cell lines BW5147 and EL4, as well as 70Z/3 and MEL. 134 putative B and T cell-specific cDNA clones and 58 putative B cell-specific cDNA clones were isolated. Elimination of cDNA clones with inserts smaller than ~300 bp, identification of lymphoid-specific cDNA clones by RNA blot analysis (see below), and elimination of redundant clones by probing bacteriophage lifts or dot blots of bacteriophage DNA with nick-translated inserts of cDNA clones already characterized, yielded fourteen unique lymphoid-specific cDNA clones. These cDNA clone inserts were subcloned into Bluescript (Stratagene) and the ends were sequenced using T3 and T7 primers and Sequenase (USB). Seven of these fourteen cDNAs were found to represent previously identified genes by nucleotide sequence comparison with sequences in GenBank.

RNA analysis

Total cytoplasmic RNA was prepared from cell lines and poly(A)$^+$ RNA was selected by passage over oligo(dT)-cellulose (BMB) (Sambrook et al., 1989). poly(A)$^+$ RNA samples (0.5 μg) were separated on 1.0% agarose gels containing 2.2 M formaldehyde, transfered to Hybond-N nylon membranes (Amersham), UV-immobilized and probed with nick-translated inserts of the various cDNA clones. Hybridization and washing conditions are as described (Travis et al., 1991a).

DNA-cellulose binding analysis

Polypeptides derived from the cDNA clones were synthesized in a rabbit reticulocyte lysate translation system (Promega) in the presence of [^{35}S]methionine using RNA generated from linearized Bluescript subclones using bacteriophage T3 or T7 RNA polymerase. Translation of the transcripts dirived from the AL1 and GL1 cDNAs was initiated at a synthetic in-frame AUG derived from vector sequences upstream of the cDNA insert (see Results).

The [^{35}S]-labelled translation reaction products from the various cDNAs were applied to DNA-cellulose columns (10-20 μl bed volume), washed extensively with 50 mM Tris pH 7.6, 1 mM EDTA, 10% glycerol, 0.5 mM DTT, and 50 mM KCl, and then eluted in steps of increasing KCl (150, 250, and 700 mM). The eluted material was collected in fractions and 35S counts per minute determined. For the DNA cellulose binding data expressed in Figure 5A, total TCA precipitable counts were determined for each translation reaction product (Harlow and Lane, 1988) and amounts of eluted protein expressed as percentages of this value. For the DNA cellulose binding data in Figure 5B, an aliquat of each translation reaction product was resolved by SDS gel electrophoresis and autoradiography using Amplify (Amersham), to allow for a comparison of the amount of each protein applied to the columns.

Primer extension analysis

Primer extension analysis of endogenous LEF-1 and DL4 transcripts was performed using 1μg of poly(A)$^+$ RNA from NIH3T3 fibroblasts or 70Z/3 pre-B cells and ^{32}P end-labelled non-coding stand oligonucleotide primers and murine leukemia virus reverse transcriptase (BRL), using previously described methods (McKnight and Kingsbury, 1982). Primers used to detect LEF-1 transcripts were derived from the 5' end of the GN8 cDNA sequence (Travis et al., 1991): G(+81) is 5'-CGAGTTTCTGCCTGGCTCGCC and G(+29) is 5'-GAGCTGCCGGTGGCCCCTCGG. Primers used to detect DL4 transcripts are dirived from the 5' end of the DL4 cDNA sequence (A.T. and R.G., unpublished data): D(+110) is 5'-GGTCTGCCCCAGAGGCCAGTG and D(+66) is 5'-CCAAATCTGAGGGTACCAGCA.

Isolation of the LEF-1 promoter

Genomic clones of the LEF-1 gene were isolated from a BALB/c mouse genomic library in the bacteriophage λ vector EMBL-3 SP6/T7 (Clontech) using a ^{32}P-labelled 1.6 kb NaeI/NheI fragment of the GN8 cDNA (Travis et al., 1991a). A 2 kb XhoI fragment containing the LEF-1 promoter was identified using the G(+29) oligonucleotide (see above), end-labelled with ^{32}P. This XhoI fragment was subcloned into Bluescript (Stratagene) and sequenced using the G(+29) oligonucleotide and Sequenase (USB).

RESULTS

Isolation of novel lymphoid-specific cDNA clones

To identify novel lymphoid-specific genes, we isolated several lymphoid-specific cDNA clones from the murine pre-B cell line 70Z/3 by differential library

screening using first-strand cDNA probes generated from either 70Z/3 or murine erythroleukemia (MEL) poly(A)$^+$ RNA. The screen yielded fourteen unique lymphoid-specific cDNA clones (Table 1) (see Materials and Methods for details on the screening procedure). These fourteen clones were further characterized using RNA blot analysis and DNA sequencing to determine the expression patterns and identify novel clones.

To examine the expression patterns of the genes represented by the fourteen cDNA clones, RNA from a variety of lymphoid and non-lymphoid cell lines was examined for the presence of transcripts hybridizing to probes generated from the cDNA clone inserts. The genes could be divided into three groups based on their expression patterns: B cell-specific, B and T cell-specific, and B and T lymphoid- and myeloid-specific. Within each group, the exact expression patterns of individual genes differed.

Within the B cell-specific gene group (Figure 1A and B; Table 1), XB1 transcripts were detected only in the bone marrow-derived pre-B cell lines PD31 and 70Z/3, BL2 transcripts were detected in all pre-B cell lines, whether derived from fetal liver (HAFTL and 40E1) or from adult bone marrow (PD31 and 70Z/3), CL2 transcripts were detected in all pre-B cell lines and the B cell lines WEHI231 ans BCL1, and AL1 and FB1 transcripts were detected in all B cell lines examined, including pre-B, B, and the plasmacytoma cell line SP2. A more extensive analysis is shown for FB1 (Figure 1B), including the additional pre-B cell lines 38B9 and PD36, the additional B cell line M12, and the additional plasmacytoma cell line J558L. No transcripts were detected from any of these genes in T cells (BW5147 and EL4) or non-lymphoid cells, including fibroblast (NIH3T3), erythroid (MEL), and myeloid (WEHI3) representatives.

Within the group of B and T cell-specific genes (Figure 2; Table 1), GL1 transcripts were detected in all pre-B and T cells examine (see Chapter Two), DL4

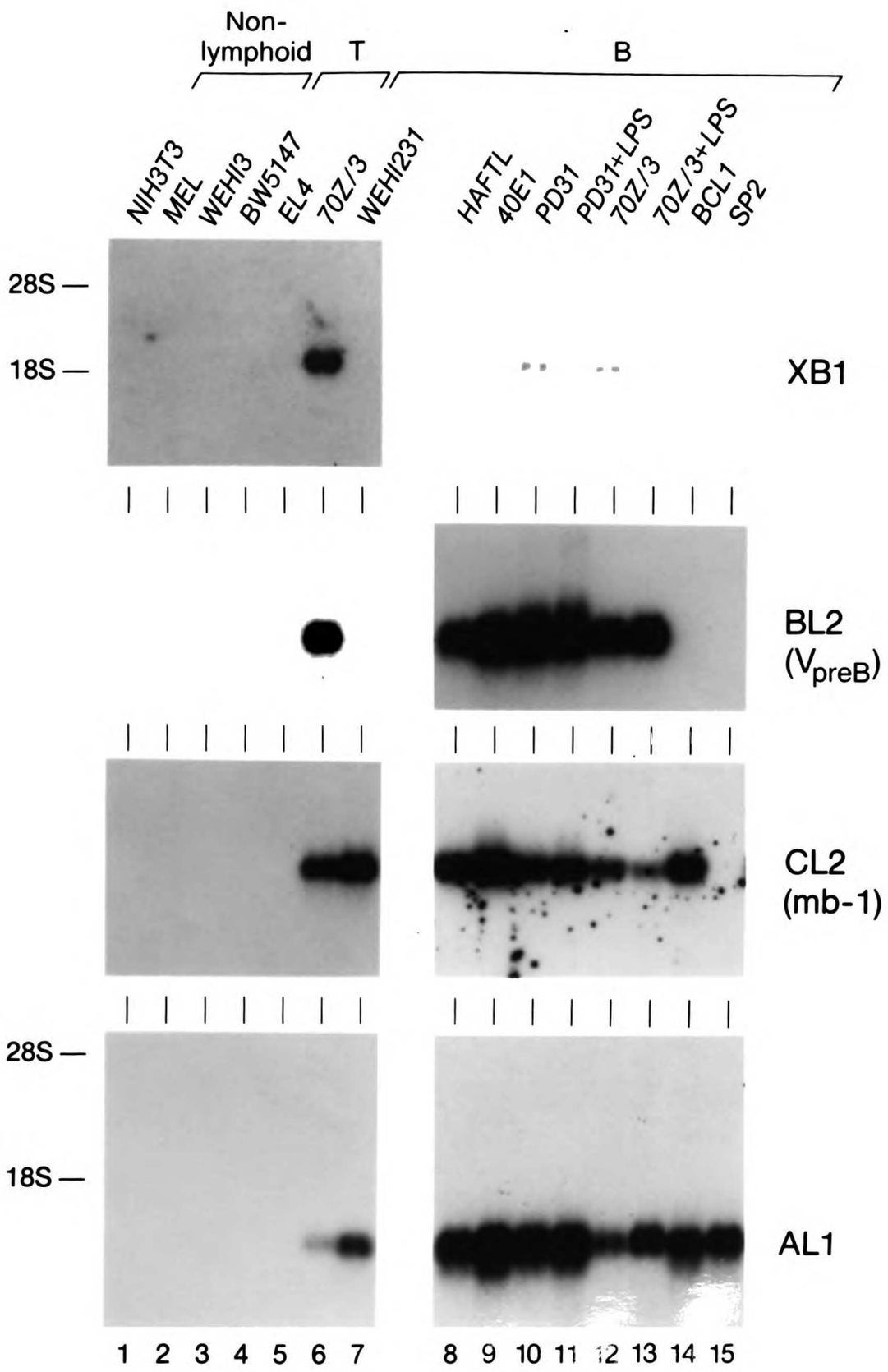
Table 1. Characterization of lymphoid-specific cDNA clones. The name, identity (if known), number of cDNA clones isolated, mRNA size, and expression pattern are summarized for each of the fourteen genes for which cDNA clones were isolated. BL2 is identical to V_{preB} (Kudo and Melchers, 1987), NL8 to λ_5 (Sakaguchi et al., 1986), CL2 to mb-1 (Sakaguchi et al., 1988), XB3 to CD19 (Tedder and Issacs, 1989), LB4 to B29 (Hermanson et al., 1988), GL1 to LEF-1/TCF-1α (Travis el al., 1991a; Waterman et al., 1991), HL4 to Ig μ heavy chain, and DL3 to Ly5 (see Thomas, 1989).

Table 1. Characterization of lymphoid-specific cDNA clones.

cDNA clone	Identity	No. of cDNA clones isolated	mRNA size (kilobases)	Cell type distribution
XB1		1	2.2	bone marrow pre-B
BL2	V_{preB}	3*	0.9	pre-B
NL8	λ_5	1*	1.2	pre-B
CL2	mb-1	1*	1.0	pre-B/B
XB3	CD19	1	2.4	pre-B/B
AL1		1*	1.3	pre-B/B/plasma
FB1		2	2.2	pre-B/B/plasma
LB4	B29	5	1.4	pre-B/B/plasma
GL1	LEF-1	3	4.2, 3.7, 2.7	pre-B/T
HL4	Ig μ	35*	1.9-3.0	pre-B/B/plasma/T
DL4		3	1.2	pre-B/B/plasma/T
LL6		1	3.1	pre-B/B/plasma/T
AL3		1	1.2	pre-B subset/T/myeloid
DL3	Ly5	1	5.6, 6.5	pre-B/B/T/myeloid

* These numbers represent lower limits since 38 redundant clones were eliminated during secondary screens.

Figure 1. RNA blot analysis of the B cell-specific clones XB1, BL2, CL2, and AL1 (Figure 1A), and FB1 (Figure 1B). Poly(A)$^+$ RNA (0.5μg) from various mouse cell lines was size-fractionated by gel electrophoresis, transfered to nylon membrane, and hybridized with nick-translated DNA probes derived from the cDNA clone inserts. The positions of the 18S and 28S rRNAs used as markers for determining mRNA sizes (Table 1) are indicated. Non-lymphoid cell lines include fibroblasts (NIH3T3), erythroleukemia cells (MEL), and myeloid cells (WEHI3). Lymphoid cell lines include T cells (BW5147 and EL4), fetal liver-derived pre-B cells (HAFTL, 40E1, and 38B9), adult bone marrow-derived pre-B cells (PD36, PD31, and 70Z/3), B cells (M12, WEHI231, and BCL1), and plasmacytomas (SP2 and J558L).


Non-lymphoid
T
B
NIH3T3
MEL
WEHI3
BW5147
EL4
70Z/3
WEHI231
HAFTL
40E1
PD31
PD31+LPS
70Z/3
70Z/3+LPS
BCL1
SP2
28S
18S
XB1
BL2
(VpreB)
CL2
(mb-1)
28S
18S
AL1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

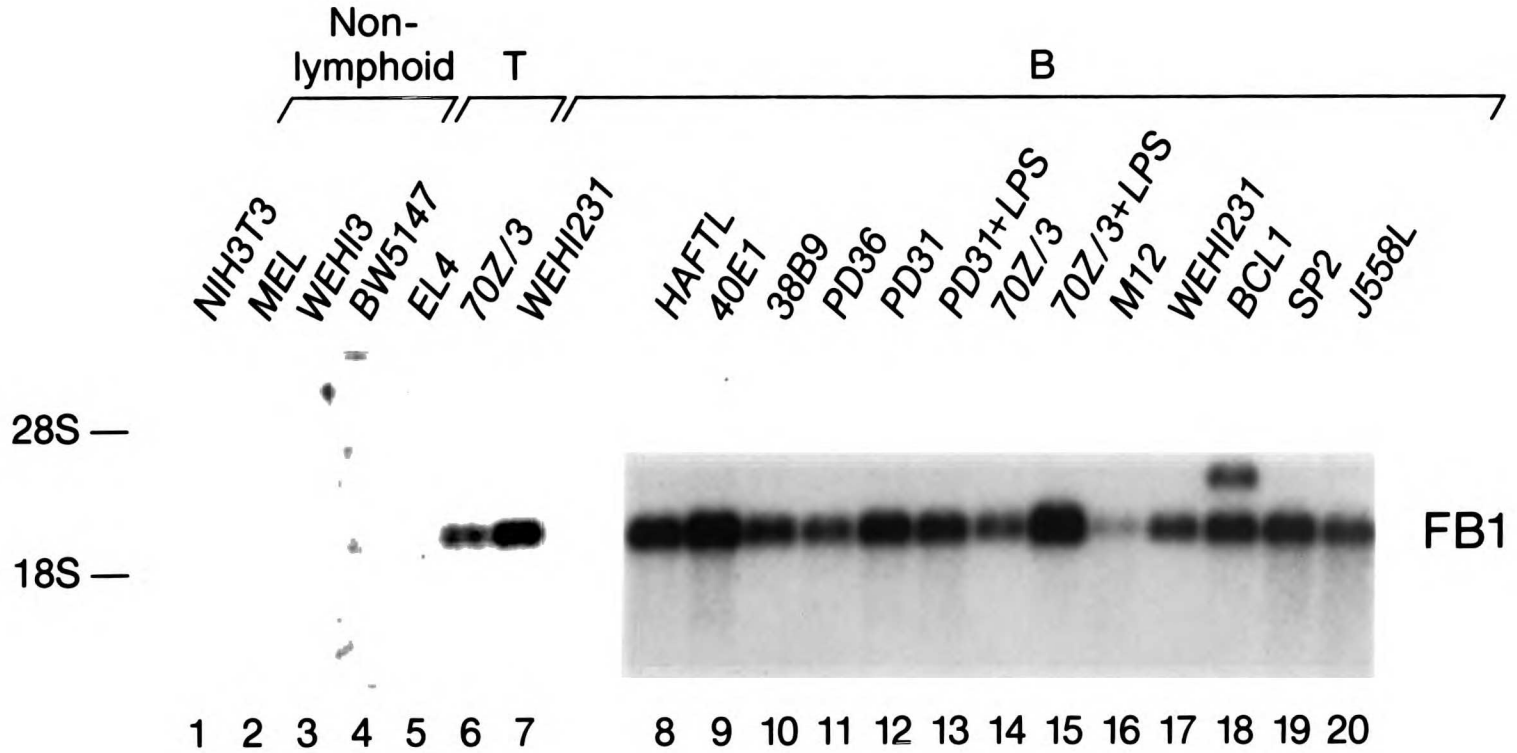
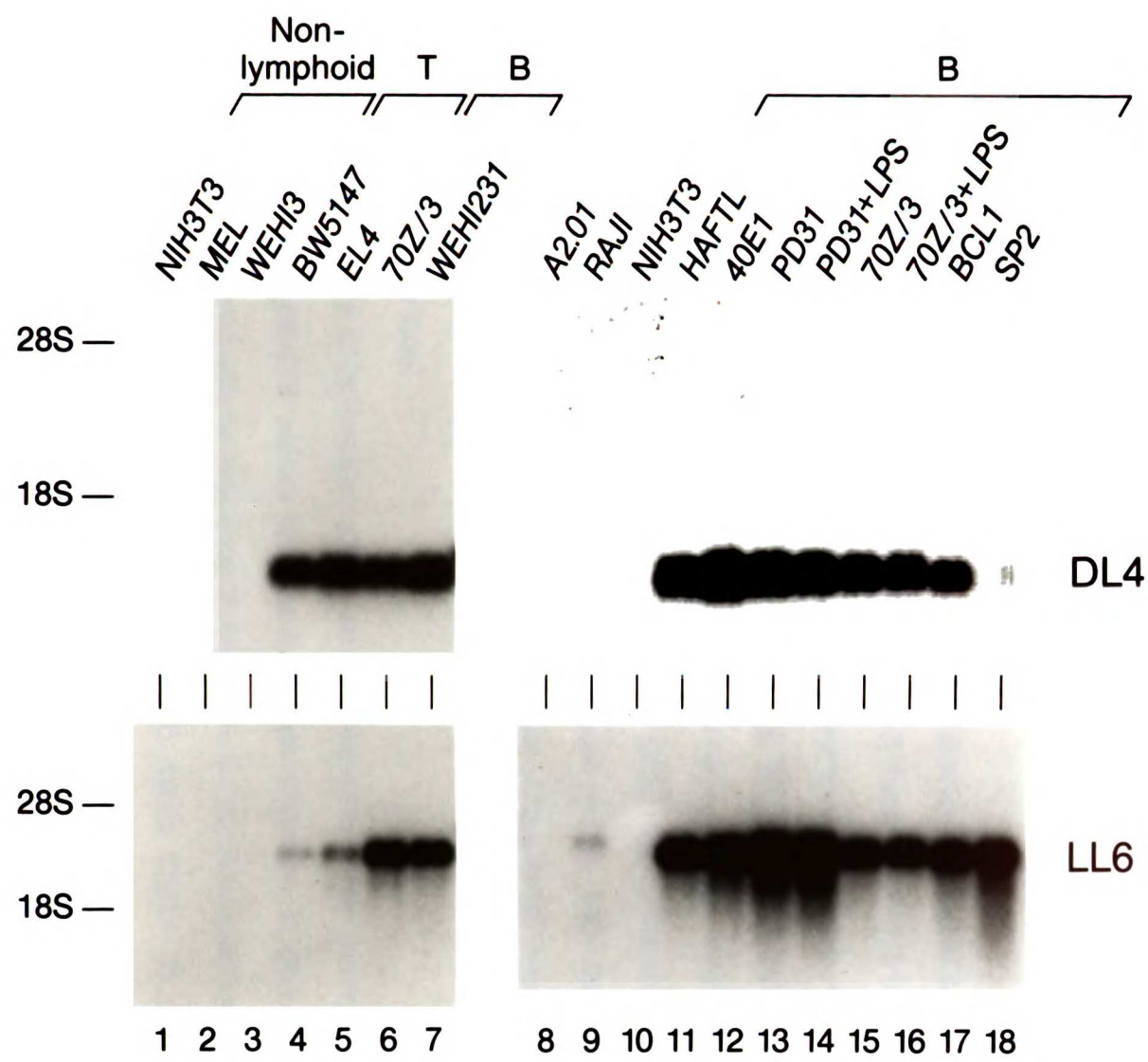

Non-
lymphoid
T
B
NIH3T3
MEL
WEHI3
BW5147
EL4
70Z/3
WEHI231
HAFTL
40E1
38B9
PD36
PD31
PD31+LPS
70Z/3
70Z/3+LPS
M12
WEHI231
BCL1
SP2
J558L
28S —
18S —
FB1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Figure 2. RNA blot analysis of the B and T cell-specific clones DL4 and LL6. RNA from various cell lines was examined as described in the legend to Figure 1 using DNA probes generated from the DL4 and LL6 cDNA clone inserts. A2.01 and RAJI are human lymphoid T and B cell lines, respectively. The other cell lines are described in the legend to Figure 1. For RNA blot analysis of the pre-B and T cell-specific clone GL1, see Chapter 2, Figure 3.


Non-lymphoid
T
B
B
NIH3T3
MEL
WEHI3
BW5147
EL4
70Z/3
WEHI231
A2.01
RAJI
NIH3T3
HAFTL
40E1
PD31
PD31+LPS
70Z/3
70Z/3+LPS
BCL1
SP2
28S
18S
DL4
28S
18S
LL6
1 2 3 4 5 6 7
8 9 10 11 12 13 14 15 16 17 18

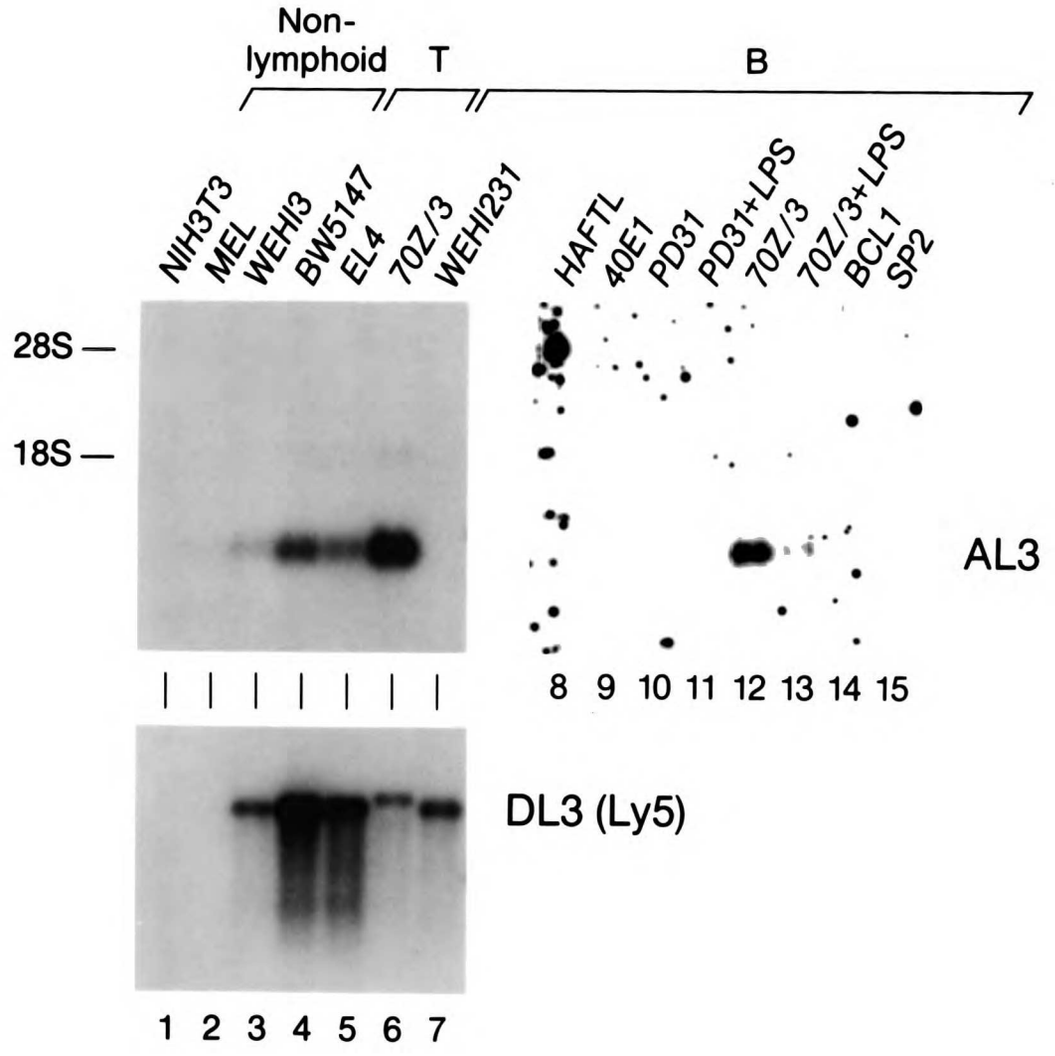
transcripts were detected in all murine B and T cells, and LL6 transcripts were detected in all B cells, including the human B cell line RAJI, and at lower levels in the murine T cell lines BW5157 and EL4, but not in the human T cell line A2.01 (Figure 2). No transcripts from these genes were detected in fibroblast (NIH3T3) or myeloid (WEHI3) representatives, but a longer exposure revealed low levels of DL4 transcripts in the erythroid (MEL) representative (data not shown) indicating that DL4 is not strictly lymphoid-specific.

Within the group of B and T lymphoid- and myeloid-specific genes (Figure 3; Table 1) AL3 transcripts were detected in the pre-B cell lines 70Z/3 and HAFTL, the T cell lines BW5147 and EL4, and at low levels in the myeloid cell line WEHI3. DL3 transcripts were detected in the pre-B (70Z/3), B (WEHI231), T (BW5147 and EL4) and myeloid (WEHI3) cell lines examined. No transcripts from either gene were detected in fibroblast (NIH3T3) or erythroid (MEL) representatives.

Maturation of the pre-B cell lines PD31 and 70Z/3 with bacterial lipopolysaccharide (LPS; Nelson et al., 1984) did not substantially effect the transcript levels for any of the genes except GL1, were it caused a three- to sixfold decrease (see Chapter Two).

To determine which of the fourteen cDNA clones represented known genes, they were subcloned into plasmid vectors and the nucleotide sequence of their ends determined and compared with the sequences in GenBank. Seven of the cDNA clones had sequence identity with known lymphoid-specific genes (Table 1). The remaining seven represented novel genes, although GL1 has now been extensively characterized and termed LEF-1 (see Chapter Two) and is the mouse homolog of TCF-1α (Waterman et al., 1991). The transcript sizes, expression patterns, and number of cDNA clones isolated for each gene are summarized in Table 1.

Figure 3. RNA blot analysis of the B and T lymphoid- and myeloid-specific clones AL3 and DL3. RNA from the various cell lines was examined as described in the legend to Figure 1 using DNA probes derived from the AL3 and DL3 cDNA clone inserts. The cell lines are described in the legend to Figure 1.

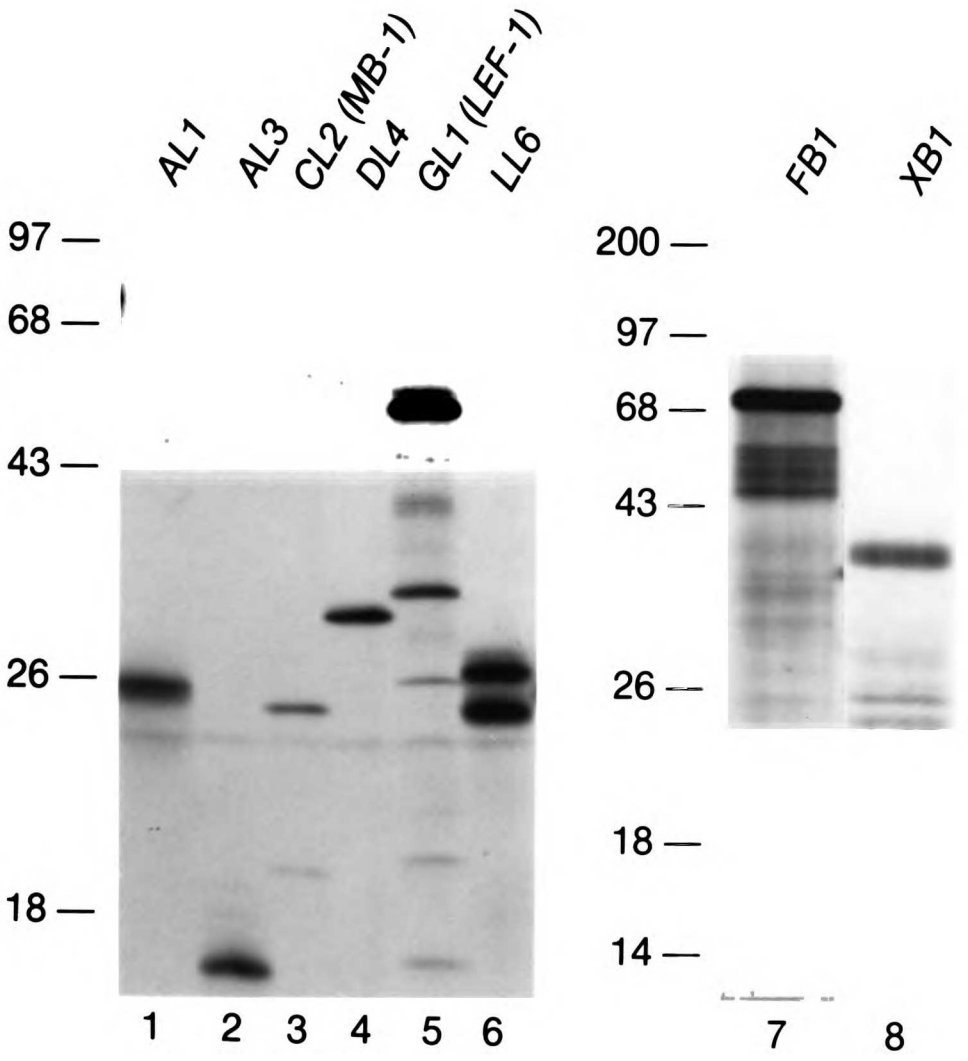

Non-
lymphoid
T
B
NIH3T3
MEL
WEHI3
BW5147
EL4
70Z/3
WEHI231
HAFTL
40E1
PD31
PD31+LPS
70Z/3
70Z/3+LPS
BCL1
SP2
28S —
18S —
AL3
8 9 10 11 12 13 14 15
DL3 (Ly5)
1 2 3 4 5 6 7

Identification of cDNAs encoding putative DNA binding proteins

To examine polypeptides encoded by the seven novel cDNA clones for the ability to bind to DNA-cellulose, the cDNA clone inserts were subcloned into vectors containing an ATG in each of the three reading frames, transcribed with T7 RNA polymerase, and the transcripts were translated in a reticulocyte lysate in the presence of [^{35}S]methionine. The translation products were examined by SDS gel electrophoresis to identify the reading frame which produce the largest polypeptide (data not shown). Two of the cDNA clones (AL1 and GL1) were found to produced a larger polypeptide with the artificial ATG added. The remaining five (AL3, DL4, LL6, FB1, and XB1) appear to be full length or posess an in-frame ATG close to their 5' ends. The translation products from the seven novel cDNA clones (and the CL2 clone which represents the mb-1 gene (Sakaguchi et al., 1988)) are shown in Figure 4.

Two of the seven novel cDNA clones, GL1 and FB1, yielded proteins that bound to DNA-cellulose (Figure 5). Binding of the GL1-derived protein was compared to that of the glucocorticoid receptor translated under the same conditions using RNA derived from the T3.1118 cDNA subclone (Miesfeld et al., 1986) (Figure 5A). Approximately 13% of the GL1-derived protein remained bound to the DNA-cellulose column after extensive washing with 50 mM KCl and was eluted in the 250 and 700 mM KCl fractions, whereas only 2% of the T3.1118-derived protein remained bound after washing and eluted in the 150 mM KCl fraction. Binding of the FB1-derived protein was less extensive than that of the GL1-derived protein, but greater than that of proteins derived from brome mosaic virus (BMV) RNA used as a negative control (Figure 5B). A protein gel autoradiogram of BMV, GL1, and FB1 translation products is shown as an inset to Figure 5B for comparison of the amount of each protein applied to the DNA-cellulose columns. The FB1-derived protein

Figure 4. Proteins derived from the lymphoid-specific cDNA clones AL1, AL3, CL2, DL4, GL1, LL6, FB1, and XB1. [^{35}S]methionine-labelled polypeptides were synthesized in vitro and size-fractionated by SDS-PAGE. Reticulocyte lysate was incubated with T7 transcripts generated from subclones of the AL1 or GL1 cDNAs, or with T3 transcripts generated from subclones of the AL3, CL2, DL4, LL6, FB1, or XB1 cDNAs. Numbers to the left of lanes 1-6 and 7-8h indicate the size (in kilodaltons) and position of the molecular weight markers. Inferred molecular weights of each polypeptide are shown in the table on the right.

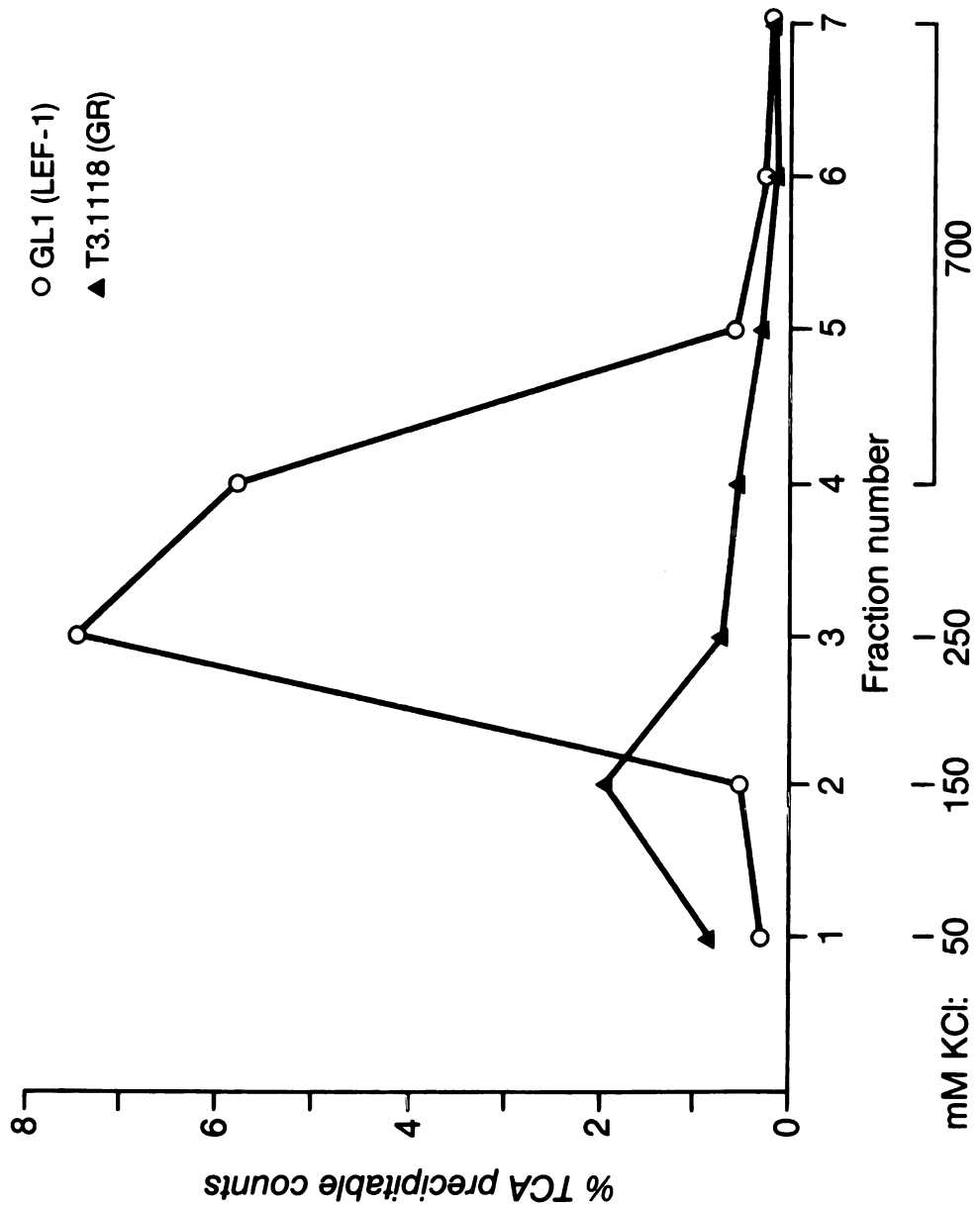


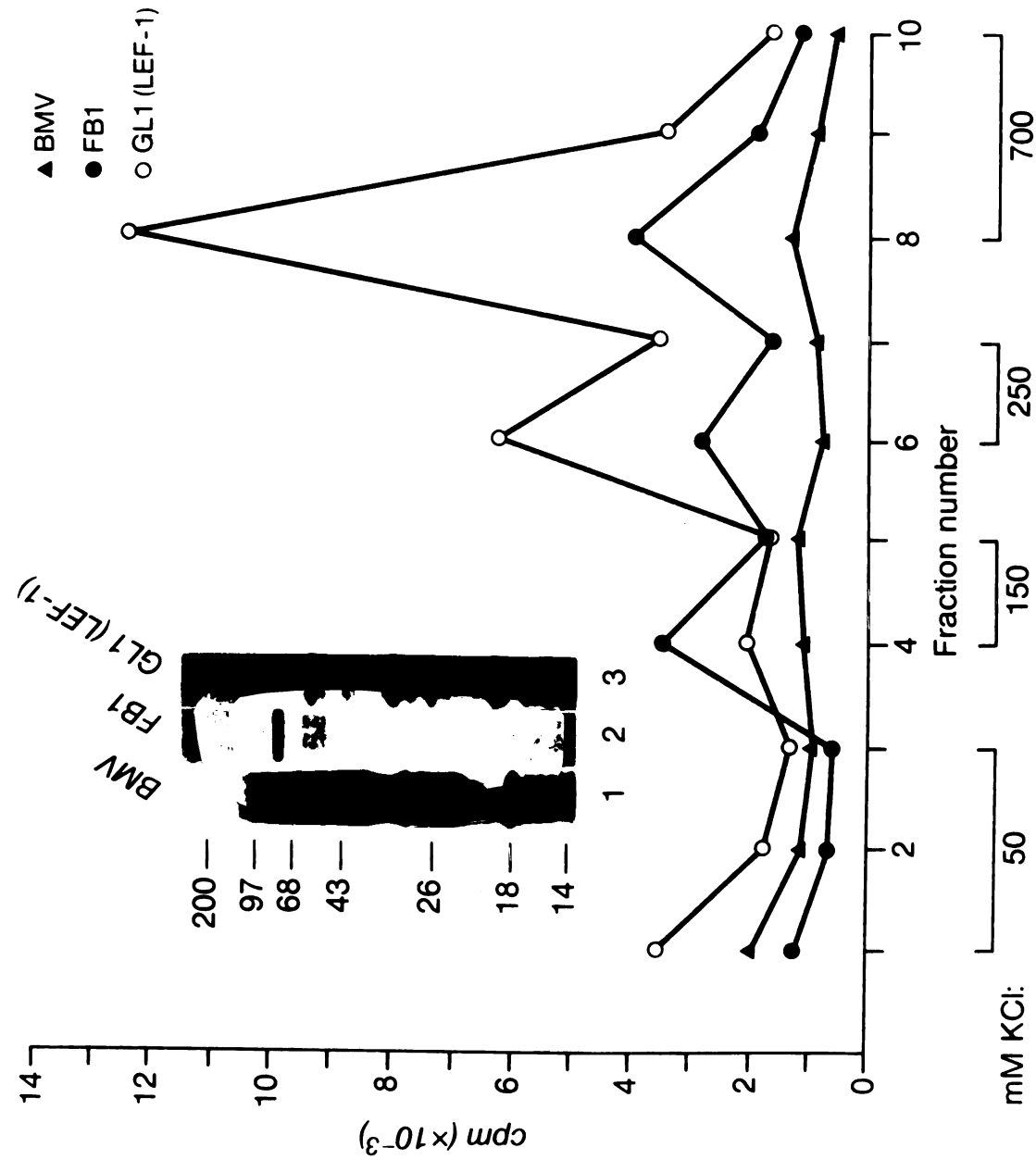
Translated cDNA	MW
AL1	26,000
AL3	16,000
CL2 (MB-1)	25,000
DL4	31,000
GL1 (LEF-1)	53,000
LL6	28,000
FB1	76,000
XB1	38,000

Figure 5. DNA-cellulose binding properties of the proteins derived from the GL1 and FB1 lymphoid-specific cDNA clones.

(A) The GL1-derived protein binds to DNA-cellulose more extensively than the glucocorticoid receptor. [^{35}S]methionine-labelled polypeptides were synthesized in vitro by incubating reticulocyte lysate with T7 transcripts from the GL1 cDNA subclone or SP6 transcripts from the T3.1118 subclone of a glucocorticoid receptor cDNA (Miesfeld et al.,1986). The translation products were applied to DNA-cellulose columns, washed to remove unincorperated [^{35}S]methionine, and eluted with a step gradient of increasing KCl. The x-axis is the fraction number, with the KCl concentration used to elute each fraction shown below. The y-axis is the percent of the TCA-precipitable ^{35}S counts applied to the column which eluted in each fraction.

(B) The FB1-derived protein binds to DNA-cellulose. [^{35}S]methionine-labelled polypeptides, synthesized by incubating reticulocyte lysate with brome mosaic virus (BMV) transcripts, T3 transcripts from the FB1 cDNA subclone, or T7 transcripts from the GL1 cDNA subclone, were applied to DNA-cellulose columns, washed to remove unincorperated [^{35}S]methionine, and eluted with a step gradient of increasing KCl. The y-axis is the ^{35}S counts per minute ($x10^{-3}$) of each fraction. An autoradiogram of an SDS gel of the BMV-, FB1-, and GL1-derived ^{35}S-labelled proteins is shown as an inset for comparison of the amounts of each protein applied to the DNA-cellulose columns.


○ GL1 (LEF-1)
▲ T3.1118 (GR)
% TCA precipitable counts
0
2
4
6
8
mM KCl:
50
150
250
700
1
2
3
4
5
6
7
Fraction number

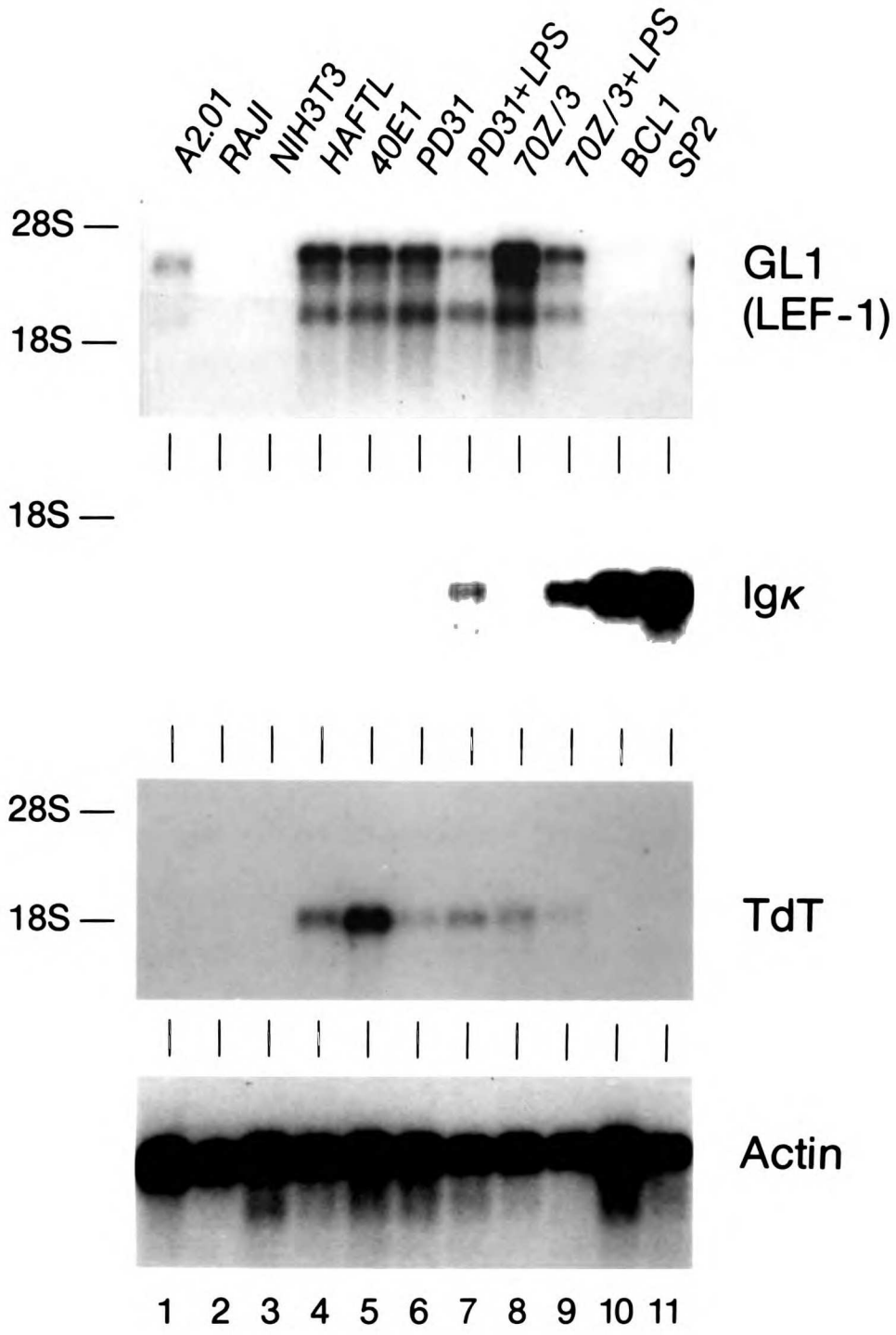

▲ BMV
● FB1
○ GL1 (LEF-1)
cpm (×10⁻³)
14
12
10
8
6
4
2
0
2
4
6
8
10
Fraction number
mM KCl:
50
150
250
700
BMV
FB1
GL1 (LEF-1)
1 2 3
200
97
68
43
26
18
14

eluted at all three KCl concentration steps (150, 250 and 700 mM), whereas GL1-dirived protein eluted only at the two higher concentartions (250 and 700 mM).

Comparison of the expression pattern of GL1 (LEF-1) with that of other lymphoid-specific genes

Since the GL1 cDNA clone represents a putative lymphoid-specific regulator and shuts off during B cell differentiation (see Chapter Two), the expression pattern of GL1 was compared with those of two other lymphoid-specific genes regulated during B cell differentiation (Figure 6). Within the B cell lineage, GL1 transcripts were detected only in pre-B cells (Figure 6, top panel, lanes 4-9) and upon differentiation of PD31 and 70Z/3 pre-B cells with LPS, transcript levels decreased three- to sixfold (compare lanes 7 and 9 with 6 and 8). Transcripts from the light chain immunoglobulin κ (Ig κ) gene (Figure 6, second panel) were detected in LPS treated PD31 and 70Z/3 pre-B cells, the B cell line BCL1, and the plasmacytoma cell line SP2. No Ig κ transcripts were detected in the pre-B cell lines HAFTL, 40E1, and 70Z/3, and only low levels were detected in untreated PD31 pre-B cells. Thus the Ig κ gene has a reciprical expression pattern from the LEF-1 gene during B cell differentiation. Transcripts from the terminal deoxynucleotidyl transferase (TdT) gene (Figure 6, third panel) were detected in pre-B cells only. Among the pre-B cell lines, 40E1 (lane 5) contained the most TdT transcripts, and LPS treatment increased transcript levels twofold in PD31 and decreased transcript levels fourfold in 70Z/3. Thus the expression pattern of the TdT gene parallels that of the LEF-1 gene in that it is pre-B [and T (Landau et al., 1984)] cell-specific, but differs substantially in expression level within various pre-B cell lines (Figure 6, compare first and third panels).

Figure 6. RNA blot analysis to compare the expression pattern of GL1 (LEF-1) with those of immunoglobulin κ (Igκ), and terminal deoxynucleotidyl transferase (TdT). RNA samples from the various cell lines were examined as described in the legend to Figure 1 using DNA probes generated from the GL1 cDNA clone insert, the constant region of the Igk gene, or a cDNA clone of the TdT gene (Landau et al., 1984). A rodent actin probe was used to control for RNA levels. The cell lines are described in the legends to Figures 1 and 2.


A2.01
RAJI
NIH3T3
HAFTL
40E1
PD31
PD31+LPS
70Z/3
70Z/3+LPS
BCL1
SP2
28S
18S
GL1
(LEF-1)
18S
Igκ
28S
18S
TdT
Actin
1 2 3 4 5 6 7 8 9 10 11

The 2.7 kb LEF-1 transcript does not contain the 5' untranslated region present in the 3.7 and 4.2 kb transcripts

When RNA from murine pre-B cell lines was examined for the presence of LEF-1 transcripts using a labelled restriction fragment from the open reading frame of the LEF-1 cDNA as the probe, three transcripts of 2.7, 3.7, and 4.2 kb were detected (Figure 7, top panel). When a duplicate RNA blot was probed with a labelled restriction fragment comprising the 5' untranslated region of the GN8 cDNA of LEF-1 (see Chapter Two, Figure 1A), only the 3.7 and 4.2 kb transcripts were detected (Figure 7, bottom panel). Thus, the 2.7 kb LEF-1 transcript lacks the 5' untranslated region present in the 3.7 and 4.2 kb LEF-1 transcripts.

A putative LEF-1 promoter lacks a consensus TATA element and appears to initiate transcription from multiple sites

The 5' ends of endogenous LEF-1 transcripts were mapped using a primer extension assay (Figure 8A). Two oligonucleotide primers comprising non-coding strand sequences from the 5' end of the GN8 cDNA of LEF-1 (see Chapter Two, Figure 1A) were used. The G(+81) primer (lanes 2 and 3) comprises GN8 sequences from +81 to +61, and the G(+29) primer (lanes 4 and 5) comprises GN8 sequences from +29 to +9 (cDNA coordinates as in Chapter Two, Figure 1b). A set of primer extension products generated with the G(+81) primer using RNA from 70Z/3 pre-B cells, but not using RNA from NIH3T3 fibroblasts (compare lanes 2 and 3) appear to represent multiple 5' ends of LEF-1 transcripts. A similar set of primer extension products mapping to the same region were generated with the G(+29) primer and 70Z/3 RNA, but not with NIH3T3 RNA (compare lanes 4 and 5). A set of weaker bands generated with both 70Z/3 and NIH3T3 RNA using the G(+29) primer appear to represent artifactual primer extension products. The fact that both primers generate multiple extension products that map to the same

Figure 7. RNA blot analysis demonstrating that the 2.7 kb LEF-1 transcript lacks the 5' untranslated region present in the 3.7 and 4.2 kb transcripts. Duplicate blots of RNA from the various cell lines were probed with nick-translated HincII fragments comprising either a 0.36 kb segment of the open reading frame (position 1158 to 1517) of the LEF-1 cDNA (upper panel), or a 1.16 kb segment (position 1 to 1157) containing the 5' untranslated region of the LEF-1 cDNA (lower panel). The cell lines are described in the legends to Figures 1 and 2.

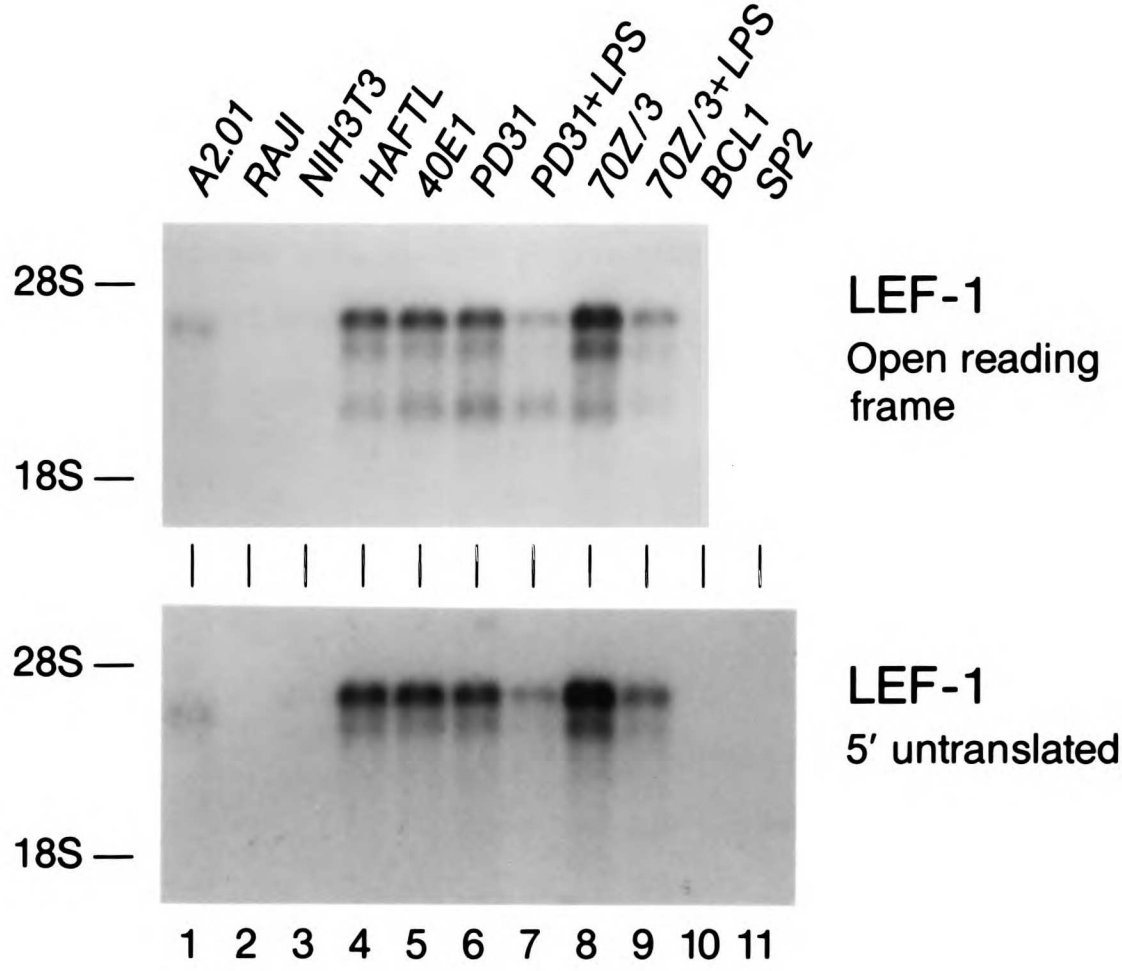

A2.01
RAJI
NIH3T3
HAFTL
40E1
PD31
PD31+LPS
70Z/3
70Z/3+LPS
BCL1
SP2
28S —
18S —
LEF-1
Open reading frame
28S —
18S —
LEF-1
5′ untranslated
1 2 3 4 5 6 7 8 9 10 11

Figure 8. A putative promoter of the LEF-1 gene.

(A) Primer extension mapping of the 5' ends of endogenous LEF-1 transcripts in RNA from 70Z/3 pre-B cells and NIH3T3 fibroblasts. The "G(+81)" and "G(+29)" oligonucleotide primers comprise non-coding strand sequences of the 5' end of the GN8 cDNA of LEF-1 from +81 to +61 and +29 to +9, respectively (cDNA coordinates as in Chapter Two, Figure 1).

(B) Nucleotide sequence of the coding strand of a putative LEF-1 gene promoter from position -145 to +118. Strong and weak putative transcription initiation sites are designated with thick and thin arrows, respectively. The +1 position is arbitrarily assigned to the strongest putative initiation site. Bracketed sequences indicate putative binding sites for the transcription factor Sp1.

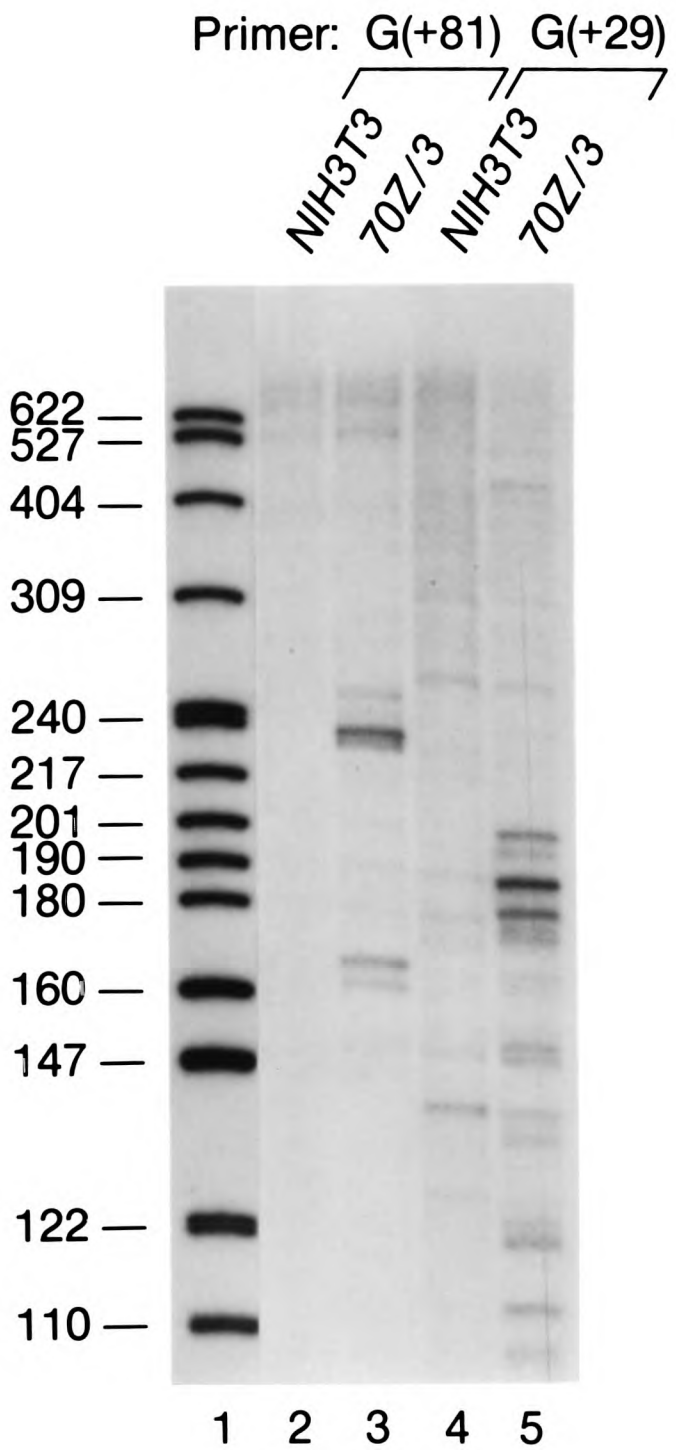
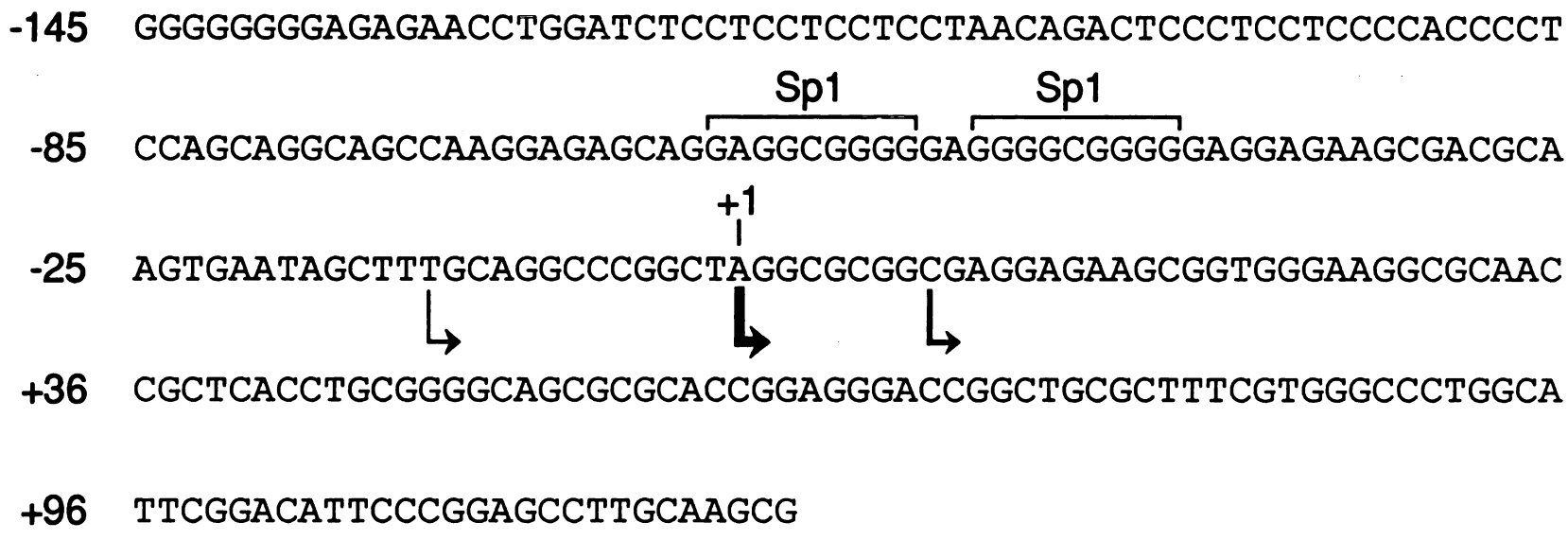

Primer: G(+81) G(+29)
NIH3T3
70Z/3
NIH3T3
70Z/3
622
527
404
309
240
217
201
190
180
160
147
122
110
1 2 3 4 5

Figure 8B


-145 GGGGGGGGAGAGAACCTGGATCTCCTCCTCCTCCTAACAGACTCCCTCCTCCCCCACCCCT
Sp1
Sp1
-85 CCAGCAGGCAGCCAAGGAGAGCAGGAGGCGGGGGAGGGGCGGGGGAGGAGAAGCGACGCA
+1
-25 AGTGAATAGCTTTGCAGGCCCGGCTAGGCGCGGCGAGGAGAAGCGGTGGGAAGGCGCAAC
+36 CGCTCACCTGCGGGGCAGCGCGCACCGGAGGGACCGGCTGCGCTTTCGTGGGCCCTGGCA
+96 TTCGGACATTCCCGGAGCCTTGCAAGCG

region suggests that transcription from the LEF-1 promoter initiates from multiple sites, although it is concievable that some or all of the shorter products represent strong stops for reverse transcriptase.

The nucleotide sequence of the putative LEF-1 promoter and the positions of the RNA 5' ends mapped by primer extension are shown in Figure 8B. The sequence lacks a concensus TATA element (see Chapter Three for a discussion of TATA-less promoters). Putative binding sites for the ubiquitous transcription factor Sp1, a feature of other TATA-less promoters (see Chapter Three), are found around -46 and -57 relative to the putative major transcription initiation site.

The DL4 cDNA is close to full length and appears to encode a novel protein with a putative transmembrane domain.

The 5'-ends of endogenous DL4 transcripts were mapped by primer extension (Figure 9A). Two oligonucleotide primers comprising DL4 cDNA sequences from +110 to +90 (D(+110)) and from +66 to +46 (D(+66)) used. Primer extension products of ~127 and ~83 nucleotides were generated with the D(+110) and D(+66) primers, respectively, using RNA from 70Z/3 pre-B cells, but not from NIH3T3 fibroblasts. Both extension products map the 5' end of the DL4 transcripts to a position ~17 nucleotides upstream of the 5' end of the DL4 cDNA, suggesting that the DL4 cDNA is close to full length.

The nucleotide sequence of both strands of the DL4 cDNA was determined and an open reading frame of 244 amino acids identified (data not shown). The predicted molecular weight of this protein (26 kD) is less than that of the in vitro translation product of the DL4 cDNA (31 kD, Figure 4). No significant homology was found when the amino acid sequence of the putative DL4 protein was compared with sequences in the Protein Identification Resource/NBRF. Inspection of the sequence revealed a 24 amino acid region of hydrophobic amino acids from position 21 to 44

Figure 9. The DL4 cDNA is close to full length and encodes a putative transmembrane protein.

(A) Primer extension mapping of the 5' ends of endogenous DL4 transcripts in RNA from 70Z/3 pre-B cells and NIH3T3 fibroblasts. The "D(+110)" and "D(+66)" oligonucleotide primers comprise non-coding strand sequences of the 5' end of the DL4 cDNA from +110 to +90 and +66 to +46, respectively.

(B) Schematic diagram of the DL4 cDNA clone and longest open reading frame. The single line represents the DL4 cDNA. The open reading frame is represented above. The hatched box (TM) depicts a region of hydrophobic amino acids which comprise a putative transmembrane domain. Numbers represent amino acid coordinates.

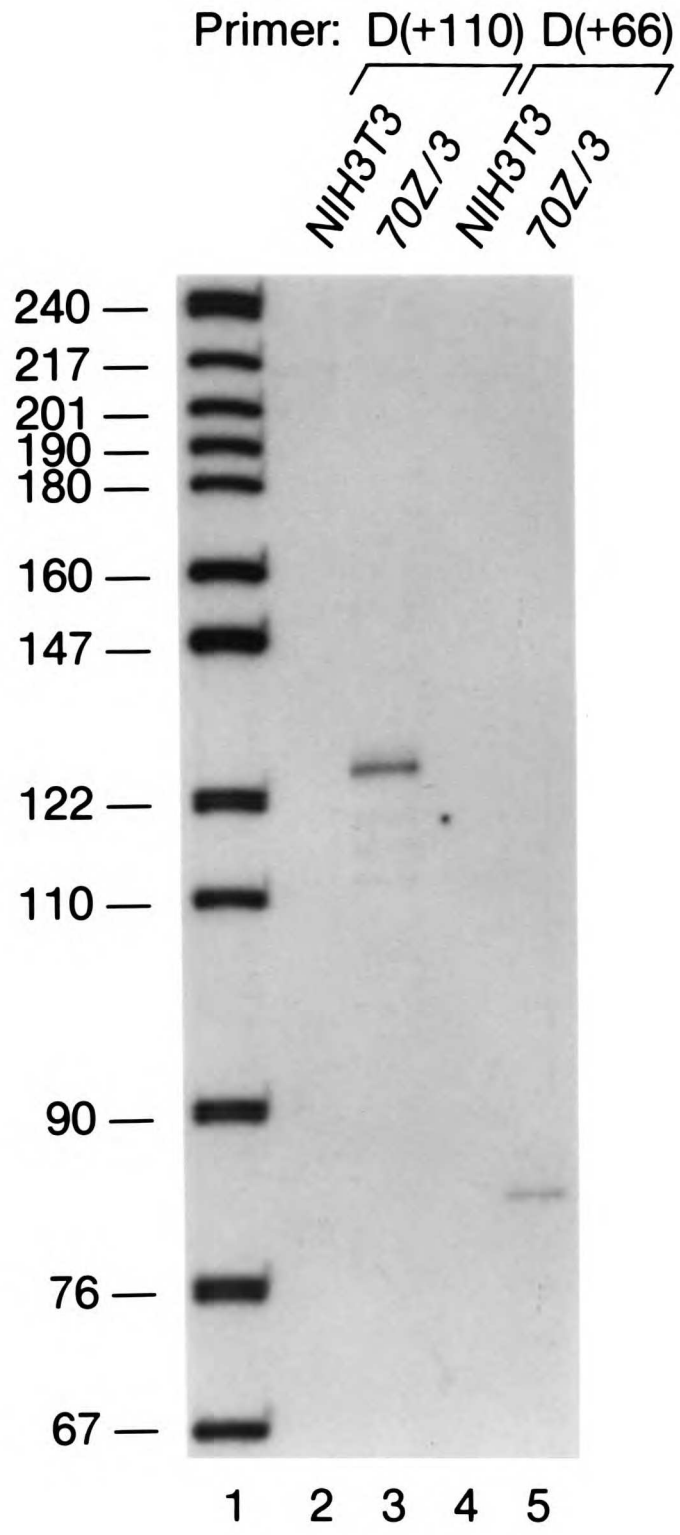
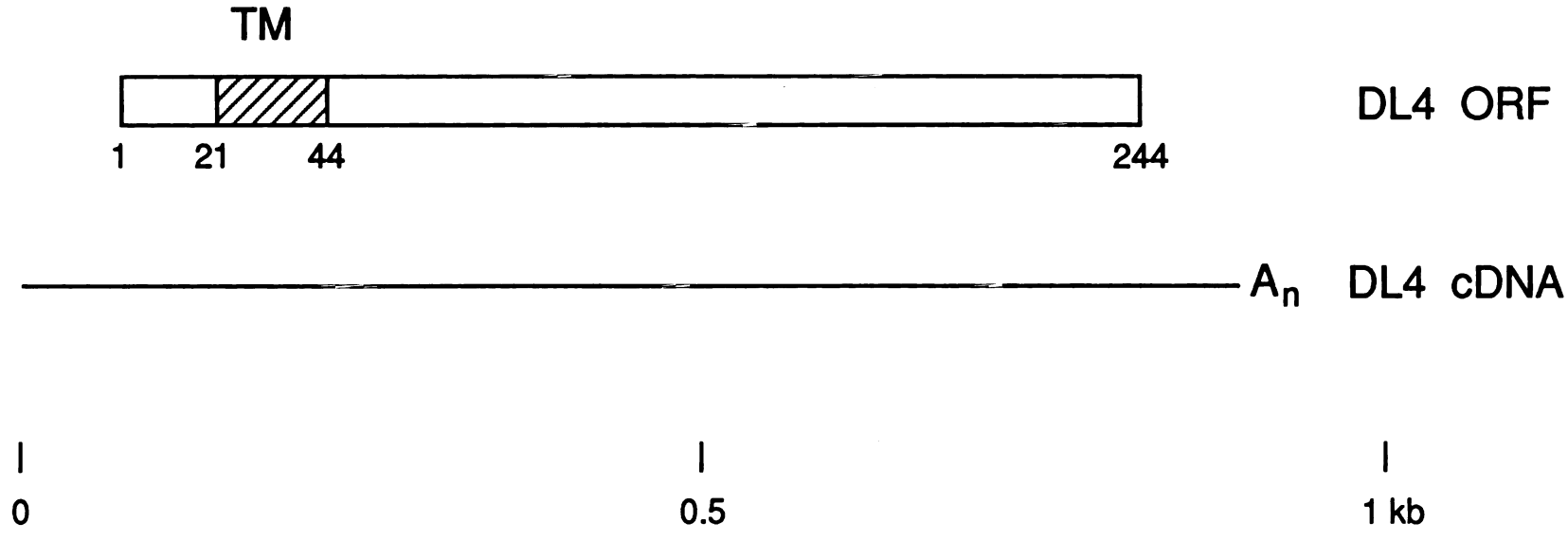

Primer: D(+110) D(+66)
NIH3T3
70Z/3
NIH3T3
70Z/3
240
217
201
190
180
160
147
122
110
90
76
67
1
2
3
4
5

Figure 9B



(Figure 9B). Thus the DL4 cDNA represents a novel lymphoid-specific gene encoding a putative transmembrane protein.

DISCUSSION

Differential cDNA library screening has been used previously to identify novel B lymphocyte-specific genes. Included are the pre-B cell-specific genes V_{preB} and λ_5 (Sakaguchi et al., 1986; Kudo and Melchers, 1987), the early B cell-specific gene mb-1 (Sakaguchi et al., 1988), and an additional seven novel lymphoid-specific genes with distinct stage-specific expression patterns in B cells or B and T cells (Yancopoulos et al., 1990).

With the aim of identifying cell type-specific regulators, we have used this method to isolate an additional seven novel lymphoid-specific cDNA clones. Two of these clones, termed GL1 and FB1, appear to encode DNA binding proteins, and thus may represent putative lymphoid-specific regulatory genes. The gene represented by GL1 is pre-B and T cell-specific and so would be expected to regulate genes expressed in either or both the B and T cell lineages. The gene represented by FB1 is expressed in all B cell stages, but not in T cells, and hence may regulate other B cell-specific genes. Further analysis of the GL1 cDNA (presented in Chapter Two) confirms that it represents a gene involved in the regulation of lymphoid-specific gene expression.

CHAPTER TWO:

LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor α enhancer function.

LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor α enhancer function

Adam Travis, Adam Amsterdam, Carole Belanger, and Rudolf Grosschedl

Howard Hughes Medical Institute and Departments of Microbiology and Biochemistry, University of California, San Francisco, California 94143-0414 USA

Lymphoid-specific cDNA clones were isolated that encode a nuclear protein with homology to the chromosomal nonhistone protein HMG-1 and to putative regulators of cell specialization, including the mammalian testis-determining factor SRY and fungal mating-type proteins. The gene represented by the isolated cDNA clones, termed LEF-1 (lymphoid enhancer-binding factor 1), is developmentally regulated and expressed in pre-B and T lymphocytes but not in later-stage B cells or nonlymphoid tissues. Both endogenous and recombinant LEF-1 were shown to bind to a functionally important site in the T-cell antigen receptor (TCR) α enhancer. Maximal TCRα enhancer activity was found to parallel the cell type-specific expression pattern of LEF-1. Moreover, forced expression of recombinant LEF-1 in late stage B cells increases TCRα enhancer function. Taken together, these data suggest that LEF-1 is a regulatory participant in lymphocyte gene expression and differentiation.

[*Key Words*: HMG box protein; TCRα enhancer; transcriptional control; lymphoid-specific gene expression]

Received February 15, 1991; revised version accepted March 20, 1991.

Lymphocyte differentiation converts progenitor cells into highly specialized cells of either the B- or the T-cell lineage. Both cell lineages involve multiple stages of differentiation that have been characterized and defined by changes in the expression patterns of their antigen receptor genes. Early-stage B lymphocytes, termed pre-B cells, express and rearrange their immunoglobulin heavy-chain (IgH) gene. Further differentiation, involving rearrangement of the light-chain locus, yields B cells that express membrane-bound immunoglobulin on their surface. Upon exposure to antigen, B cells ultimately mature into antibody-secreting plasma cells (for review, see Blackwell and Alt 1988; Cooper and Burrows 1989). Likewise, differentiating T cells undergo stage-specific changes in the expression of genes encoding the T-cell antigen receptor (TCR) and accessory molecules. Early-stage T cells undergo sequential rearrangement of their TCRβ- and TCRα-chain genes and express both CD4 and CD8 surface molecules. These cells differentiate into mature T cells expressing both TCR chains, and either the CD4 or the CD8 coreceptor molecule (for review, see Davis and Bjorkman 1988; Blackman et al. 1990).

To gain insight into some of the mechanisms underlying terminal differentiation, several studies were aimed at identifying the *cis*- and *trans*-acting regulatory components of genes specific for the differentiated phenotypes. Expression of immunoglobulin genes in B cells and TCR genes in T cells was shown previously to depend on transcriptional enhancers located 5′ or 3′ of their respective constant regions (Gillies et al. 1983; Banerji et al. 1983; Grosschedl and Baltimore 1985; Krimpenfort et al. 1988; McDougall et al. 1988; Ho and Leiden 1989; Winoto and Baltimore 1989). These enhancers function specifically in lymphocytes. The IgH enhancer is active in all B cells and in a subset of T cells (Gerster et al. 1986). In contrast, the TCRβ-chain enhancer is functional in T cells and in early-stage B cells (Krimpenfort et al. 1988; Takeda et al. 1990). Both the immunoglobulin and TCR enhancers were subsequently used to identify and purify nuclear factors that bind to specific nucleotide sequences within the enhancers. Some of these factor-binding sites were used to isolate complementary DNAs (cDNAs) of genes encoding enhancer-binding proteins (Murre et al. 1989; Henthorn et al. 1990; Ho et al. 1990; Roman et al. 1990). Surprisingly, most of the nuclear factors and genes identified by these approaches are ubiquitously expressed. To date, one gene, termed Oct-2, was isolated and shown to encode a lymphocyte-specific protein that binds to the OCTA site in both the IgH enhancer and promoter (Muller et al. 1988; Scheidereit et al. 1988; Staudt et al. 1988). The OCTA site, however, can also be recognized by other

related proteins such as Oct-1, which is expressed in virtually all cell types (Singh et al. 1986; Scholer et al. 1989).

On the basis of the notion that the binding of ubiquitous factors may obscure the identification of cell type-specific factors binding the same site, we employed an alternative strategy for the cloning of genes encoding putative lymphocyte-specific regulatory factors. We first generated a set of lymphocyte-specific cDNA clones. Binding of the encoded polypeptides to DNA was then used as a criterion to identify cDNA clones that encode putative regulatory proteins. Lymphocyte specificity is thus ensured prior to screening for the ability to bind DNA.

Here, we report the molecular cloning of cDNAs of a developmentally regulated murine gene that is expressed specifically in pre-B and T lymphocytes. This gene, termed LEF-1 (for lymphoid enhancer-binding factor 1), encodes a nuclear protein that binds to a functionally important site in the TCRα enhancer and confers maximal activity. LEF-1 is a new member of a family of regulatory proteins that share homology with the high mobility group protein 1 (HMG-1). We propose LEF-1 as a putative regulatory participant in lymphocyte differentiation.

Results

Cloning of cDNAs encoding LEF-1

To identify candidate transcriptional regulators of the lymphocyte lineage, we isolated several lymphoid-specific cDNA clones from the murine pre-B-cell line 70Z/3 by differential screening of recombinant bacteriophages with radiolabeled first-strand cDNA probes from either 70Z/3 or murine erythroleukemia (MEL) poly(A)$^+$ RNA (see Materials and methods). To examine polypeptides encoded by these cDNA clones for their ability to bind to DNA–cellulose, the inserts from eight different recombinant bacteriophages were subcloned into vectors containing an ATG in each of the three reading frames, transcribed with T7 RNA polymerase, and the transcripts were translated in a reticulocyte lysate in the presence of [^{35}S]methionine. One lymphoid-specific cDNA clone, termed GL1, yielded a protein that bound to DNA–cellulose, remained bound at 150 mM KCl, and could be eluted at 250 mM KCl (data not shown). The nucleotide sequence of the GL1 cDNA was determined and found to contain an open reading frame fused to an upstream ATG in the vector. A longer cDNA clone, termed GN8, was isolated, which extended an additional 1 kb at the 5′ end of the GL1 open reading frame (Fig. 1a). Analysis of the nucleotide sequence of the GN8 cDNA revealed a single open reading frame of 397 amino acids, beginning with an ATG at nucleotide 990 and terminating with a stop codon at nucleotide 2181 (Fig. 1b). The open reading frame is preceded by a long untranslated region with three stop codons in the same frame, suggesting that GN8 contains the entire coding sequence. Conceptual translation of the open reading frame predicts a protein with a molecular mass of 44 kD. On the basis of its lymphoid-specific expression pattern and specific binding to transcriptional enhancer sequences (see below), we term the GN8-encoded protein LEF-1.

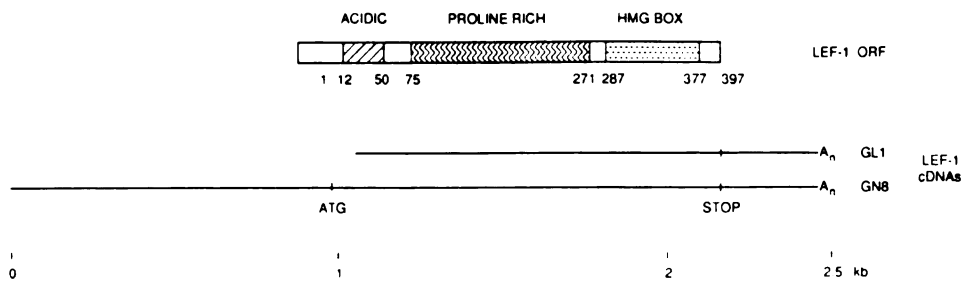
LEF-1 is an HMG box protein

The predicted amino acid sequence of LEF-1 has structural features suggestive of a transcriptional regulator. Amino acid sequence comparison of LEF-1 with sequences in the Protein Identification Resource/NBRF revealed a region of homology to the nonhistone chromosomal protein HMG1 (Wen et al. 1989) and the mating type protein Mat Mc of *Schizosaccharomyces pombe* (Kelly et al. 1988). This region of LEF-1, comprising ~85 amino acids and designated the HMG box (Fig. 1a,b), was also found to share amino acid sequence homology with the human upstream binding factor (hUBF), a nucleolar protein that binds and transcriptionally activates the rRNA gene promoter (Jantzen et al. 1990). hUBF contains four HMG boxes, one of which appears to be sufficient for DNA binding (Jantzen et al. 1990). More recently, additional members of this new family of proteins have been identified. A gene from the sex-determining region of the human and mouse Y chromosome, termed SRY, encodes a testis-specific HMG box protein that has been proposed to play a crucial role in testis development (Gubbay et al. 1990; Sinclair et al. 1990). Likewise, the genetically defined regulatory gene *mt a1*, which is involved in mating type specification of *Neurospora crassa*, encodes an HMG box protein (Staben and Yanofsky 1990).

Alignment of the HMG boxes of LEF-1 and other members of the family (Fig. 2) revealed 25% identity and 47% similarity with the Mc mating-type protein of *S. pombe* (Kelly et al. 1988) and 25% identity and 41% similarity with murine SRY (Gubbay et al. 1990; for the percentage of amino acid identity and similarity with other HMG boxes and for allowed conservative amino acid substitutions, see the legend to Fig. 2). With the exception of one amino acid gaps in Mat Mc and mt a1, no spacing changes were necessary for the optimal alignment of the HMG boxes. The HMG box of LEF-1 appears to be more closely related to those of the fungal mating-type proteins and SRY than to that of hUBF and HMG-1. The homology of LEF-1 with the various members of the HMG box proteins is concentrated in three regions within the HMG box: a block of eight neutral and hydrophobic amino acids flanked by charged residues from position 10 to 26, a region containing conserved hydrophobic and basic residues between positions 41 and 78, and a basic region at the carboxyl terminus that is most noticeable in LEF-1 (Fig. 2). We also note that the HMG box of LEF-1 differs from those of the other members of the family in the region from position 33 to 40, where amino acid conservation is exclusive of LEF-1 (Fig. 2).

The predicted amino acid sequence of LEF-1 also includes a segment of 197 amino acids containing 37 proline residues and a 39-amino-acid region containing 14 aspartic and glutamic acid residues (Fig. 1a). Proline-rich

a



b

```
1    CGGAGGAGCCGAGGGGCCACCGGCAGCTCTTTGCTTTGACCGAGGGGAGGCGGGGAGAGAGGCGAGCCAGGCAGAAACTCGAGCCGGGAACAAAGAGGGATCGGGCTCGAGTGTGTGTG
120  TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTAAGTGTGTGTATCGGCCCGAGCTGCGGGCTGGAACATTTGGGCCCGAAGCTCCTGCTGTGACTCCCCAAGACTCCGCCGTGCCAGC
239  CACCGCCGATTCCCAGCGCTCATCATCACAAACTTTATTCTTGGCAAACTTCTCTTTTTCTCCCCTCCCCCTCCAGCAGATTAAATGCTCCTCCAGAAGGAAAACCGAAGCGAAAGGGA
358  AGGAAAGAAGCTCTAACGCGGACGTCTGCAGCCCGGTGGCTCTTTATTGTTTACTCTGAAGGAAGTGGACTTTTCGGTATTTTCTGATTCTTCTCGTACCTCCGCTGGGGCAAAGGGAG
477  CCTCTTGGCCAGCTCTCCTCTTCTCAAAAACAAAACAAAACAAAACCCCAAATCACCGAAATCAGCATCCGAGGAATCCGTGGAGAGGTTTTGCACGGTGACGCACCACGGGACCCTCG
596  TTTGCACTTTGGCAAGTCGTGTCCCCTCCGGAACCGGTGTCCTCGAAGCACAGACCCCAGCAACGAGTGCGGCGTGGGGTGGCCACCCAGCCCTGGCAGCCTAGCCTAGTGCACGCGGA
715  GCGCGTAGACGCTCTCAGAGGCGCCGGCGTGCCGCCCTCGCAGTCTGGTGCCCTTCGCCCCAGGCCGGGGCGCCGGCGTCCTTCTAAGTGGGAAAGCGCGGCAATCGCAGAGGCTCCTG
834  CAGCTGCTCCAGCCGTGGGCCCCTCGCTTGCCCACCCGGCTTCCATTCACAGTCCTCCCGCACTCAGTCTTCCAACTCCCCCACCTCACATCGACCCTCCCTTCTTGCTGTCTCGGAGT

                                        1                                  10                                  20
                                      Met Pro Gln Leu Ser Gly Gly Gly Gly Gly Gly Asp Pro Glu Leu Cys Ala Thr Asp Glu
953  GCTCCCGCAGCGGAGCGGAGATTACACAGCCGCCGGG   ATG CCC CAA CTT TCC GGA GGA GGC GGC GGG GAC CCG GAA CTC TGC GCC ACC GAT GAG

                                   30                                  40                                  50
     Met Ile Pro Phe Lys Asp Glu Gly Asp Pro Gln Lys Glu Lys Ile Phe Ala Glu Ile Ser His Pro Glu Glu Glu Gly Asp Leu Ala Asp
1050 ATG ATC CCC TTC AAG GAC GAA GGC GAT CCC CAG AAG GAG AAG ATC TTC GCC GAG ATC AGT CAT CCC GAA GAG GAG GGC GAC TTA GCC GAC

                                   60                                  70                                  80
     Ile Lys Ser Ser Leu Val Asn Glu Ser Glu Ile Ile Pro Ala Ser Asn Gly His Glu Val Val Arg Gln Ala Pro Ser Ser Gln Glu Pro
1140 ATC AAG TCA TCT TTG GTT AAC GAG TCC GAA ATC ATC CCA GCC AGC AAC GGG CAT GAG GTG GTC AGA CAA GCC CCG TCC TCT CAG GAG CCC

                                   90                                 100                                 110
     Tyr His Asp Lys Ala Arg Glu His Pro Asp Glu Gly Lys His Pro Asp Gly Gly Leu Tyr Asn Lys Gly Pro Ser Tyr Ser Ser Tyr Ser
1230 TAC CAC GAC AAG GCC AGA GAA CAC CCT GAT GAA GGA AAG CAT CCA GAC GGA GGC CTG TAC AAC AAG GGA CCC TCC TAC TCC AGT TAC TCT

                                  120                                 130                                 140
     Gly Tyr Ile Met Met Pro Asn Met Asn Ser Asp Pro Tyr Met Ser Asn Gly Ser Leu Ser Pro Pro Ile Pro Arg Thr Ser Asn Lys Val
1320 GGC TAC ATA ATG ATG CCC AAT ATG AAC AGC GAC CCG TAC ATG TCA AAT GGG TCC CTT TCT CCA CCC ATC CCG AGG ACA TCA AAT AAA GTG

                                  150                                 160                                 170
     Pro Val Val Gln Pro Ser His Ala Val His Pro Leu Thr Pro Leu Ile Thr Tyr Ser Asp Glu His Phe Ser Pro Gly Ser His Pro Ser
1410 CCC GTG GTG CAG CCC TCT CAC GCG GTC CAC CCG CTC ACC CCC CTC ATC ACC TAC AGC GAC GAG CAC TTT TCT CCG GGA TCC CAC CCG TCA

                                  180                                 190                                 200
     His Ile Pro Ser Asp Val Asn Ser Lys Gln Gly Met Ser Arg His Pro Pro Ala Pro Glu Ile Pro Thr Phe Tyr Pro Leu Ser Pro Gly
1500 CAC ATC CCG TCA GAT GTC AAC TCC AAG CAA GGC ATG TCC AGA CAC CCT CCA GCT CCT GAA ATC CCC ACC TTC TAC CCC CTG TCT CCG GGC

                                  210                                 220                                 230
     Gly Val Gly Gln Ile Thr Pro Pro Ile Gly Trp Gln Gly Gln Pro Val Tyr Pro Ile Thr Gly Gly Phe Arg Gln Pro Tyr Pro Ser Ser
1590 GGC GTT GGA CAG ATC ACC CCA CCC ATT GGC TGG CAA GGT CAG CCT GTT TAT CCC ATC ACG GGT GGA TTC AGG CAA CCC TAC CCA TCC TCA

                                  240                                 250                                 260
     Leu Ser Gly Asp Thr Ser Met Ser Arg Phe Ser His His Met Ile Pro Gly Pro Pro Gly Pro His Thr Thr Gly Ile Pro His Pro Ala
1680 CTG TCA GGC GAC ACT TCC ATG TCC AGG TTT TCC CAT CAT ATG ATT CCT GGT CCC CCT GGC CCC CAC ACA ACT GGC ATC CCT CAT CCA GCT

                                  270                                 280                                 290
     Ile Val Thr Pro Gln Val Lys Gln Glu His Pro His Thr Asp Ser Asp Leu Met His Val Lys Pro Gln His Glu Gln Arg Lys Glu Gln
1770 ATT GTA ACA CCT CAG GTC AAA CAG GAG CAC CCC CAC ACG GAC AGT GAC CTA ATG CAC GTG AAG CCT CAA CAC GAA CAG AGA AAG GAG CAG

                                  300                                 310                                 320
     Glu Pro Lys Arg Pro His Ile Lys Lys Pro Leu Asn Ala Phe Met Leu Tyr Met Lys Glu Met Arg Ala Asn Val Val Ala Glu Cys Thr
1860 GAG CCC AAA AGA CCT CAT ATT AAG AAG CCT CTG AAT GCT TTC ATG TTA TAT ATG AAA GAA ATG AGA GCG AAT GTC GTA GCT GAG TGC ACG

                                  330                                 340                                 350
     Leu Lys Glu Ser Ala Ala Ile Asn Gln Ile Leu Gly Arg Arg Trp His Ala Leu Ser Arg Glu Glu Gln Ala Lys Tyr Tyr Glu Leu Ala
1950 CTA AAG GAG AGT GCA GCT ATC AAC CAG ATC CTG GGC AGA AGA TGG CAC GCC CTC TCC CGG GAA GAG CAG GCC AAA TAC TAT GAA CTA GCA

                                  360                                 370                                 380
     Arg Lys Glu Arg Gln Leu His Met Gln Leu Tyr Pro Gly Trp Ser Ala Arg Asp Asn Tyr Gly Lys Lys Lys Lys Arg Lys Arg Glu Lys
2040 CGG AAA GAG AGA CAG CTA CAC ATG CAG CTT TAT CCA GGC TGG TCA GCG CGA GAC AAT TAT GGC AAG AAG AAG AAG AGG AAG AGA GAG AAG

                                  390                 397
     Leu Gln Glu Ser Thr Ser Gly Thr Gly Pro Arg Met Thr Ala Ala Tyr Ile OP
2130 CTA CAG GAG TCG ACT TCA GGT ACA GGT CCC AGA ATG ACA GCT GCC TAC ATC TGA AACATGGTGGTAAGAGAAGCTCCTTCCCAACGTGCAAAGCCAAGGCAG

2232 CGACCCCAGGCCCTCTTCTGGAGATGGAAGCTTGTTGAAACCCCAGACTGTCTCCACAGCTTGCCCGGCTGACCCCAAGGAACACTGACAGCAACCTTACCCTGAGGTCACTGCTAGCGC

2352 TGACCCGAAGACACAGTCACTGCCACCTCTTCCTTCTGTCGTCTACTGCAAGCGCCGACTTCCAAAAAGAAAGCCGAAAACGGTTGTTTCGGAAAAAAAAAAAAAAAAAA
```

HMG box

Figure 1. (*See facing page for legend.*)

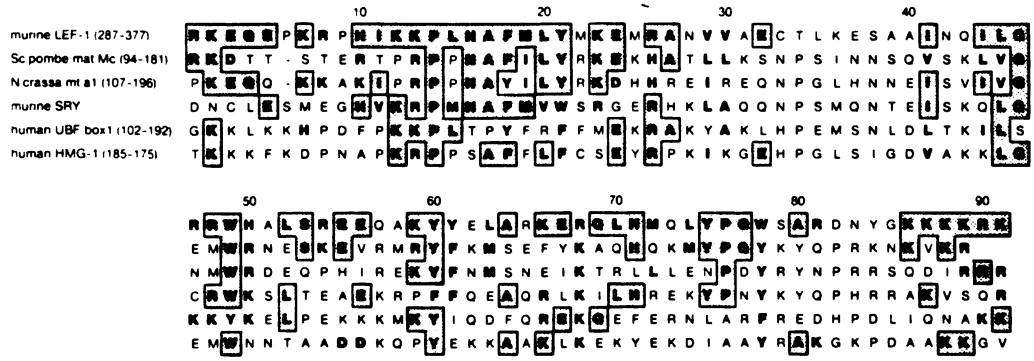


Figure 2. Amino acid sequence alignment of the HMG boxes of LEF-1 and other members of the family. The first line shows the HMG box of murine LEF-1. The amino acid positions of the HMG box in each protein are indicated in parentheses. The numbers above the amino acid sequence of LEF-1 indicate the relative positions within the HMG box. The HMG boxes of the mating-type proteins Mat Mc of *S. pombe* (Kelly et al. 1988) and mt a1 of *N. crassa* (Staben and Yanofsky 1990) show 25% and 22% amino acid sequence identity with the HMG box of LEF-1, respectively, and 47% and 41% sequence similarity. For optimal sequence alignment, one-amino-acid gaps were allowed in Mat Mc between positions 98 and 99, and in mt a1 between positions 111 and 112. The sequence identity and similarity of the HMG box of LEF-1 with that of the murine SRY (Gubbay et al. 1990) are 25% and 41%; with that of human hUBF (Jantzen et al. 1990) are 18% and 35%; and with that of human HMG-1 (Wen et al. 1989) are 21% and 32%. Amino acids identical among the various HMG boxes are boxed, and conserved amino acids are indicated in bold type. Conserved amino acid substitution groups are (K,R,H), (L,I,V), (W,Y,F), (E,D), and (L,M,I).

and acidic regions have been implicated previously in the activation functions of RNA polymerase II transcription factors (for review, see Mitchell and Tjian 1989).

Developmental expression pattern

To examine the expression pattern of LEF-1, poly(A$^+$) RNA from various mouse tissues was analyzed for the presence of transcripts hybridizing to a probe derived from the coding region of GN8 cDNA. RNA blot analysis revealed multiple-sized transcripts specifically in lymphoid tissues (lymph nodes, spleen, and thymus) and in testis (Fig. 3a). Two major transcripts of 2.7 and 4.2 kb were detected in thymus and at lower abundance in lymph nodes and spleen. Hybridization of the RNA blot with a DNA probe derived from the 5′ end of the GN8 cDNA revealed only the 4.2-kb major transcript, suggesting that the GN8 clone represents a partial cDNA of this transcript (data not shown). The 2.7-kb transcript appears to be generated by alternate promoter usage or alternate RNA processing, or may be derived from a very closely related gene. In addition, a series of low abundant transcripts ranging from 3.0 to 3.4 kb were detected in testis. Preliminary characterization of a clone isolated from a testis cDNA library using a GN8 cDNA probe indicates that these testis-specific transcripts are derived from the DNA strand opposite to that encoding the LEF-1 transcripts (Rong-guo Qiu and R. Grosschedl, unpubl.). Therefore, the LEF-1 gene appears to be expressed in a lymphoid-specific manner.

To gain further insight into the expression of LEF-1 within the lymphoid B- and T-cell lineages, RNA from murine and human cell lines representing various stages of differentiation was examined for the presence of LEF-1 transcripts (Fig. 3b). Two major LEF-1 transcripts of 4.2 and 2.7 kb and a minor LEF-1 transcript of 3.7 kb were detected in all pre-B-cell lines, whether derived from fetal liver (lanes 4–6) or from adult bone marrow (lanes 7–11). Maturation of the pre-B-cell lines PD31 and 70Z/3 with bacterial lipopolysaccharide (LPS; Nelson et al. 1984), resulted in a three- to sixfold decrease in the number of LEF-1 transcripts (cf. lane 8 with 9 and lane 10 with 11). Cell lines representing later stages of B-cell differentiation (mature B cells and plasmacytomas) did not contain any LEF-1 transcripts (lanes 12–17). However, LEF-1 transcripts were detected in all T-cell lines analyzed (lanes 18–22), irrespective of their differentiation stage. Finally, LEF-1 transcripts were not detected in any of the nonlymphoid cell lines examined (lanes 1–3), which include representatives of the myeloid (WEHI 3) and erythroid (MEL) hematopoetic lineages. Taken to-

Figure 1. Nucleotide sequence of the murine LEF-1 cDNA and predicted amino acid sequence of the encoded protein. (*a*) Schematic diagram of LEF-1 cDNA clones and the longest open reading frame. The two lines represent the cDNA clones GL1 and GN8. The open reading frame is shown above. (HMG box) The stippled box representing the region of homology with HMG-1. The hatched box depicts an acidic region (36% aspartic and glutamic acid); the zigzagged box represents a proline-rich region (19% proline). (*b*) Nucleotide sequence of the cDNA clone GN8 and predicted amino acid sequence of the encoded protein. The HMG homology is indicated by the box.

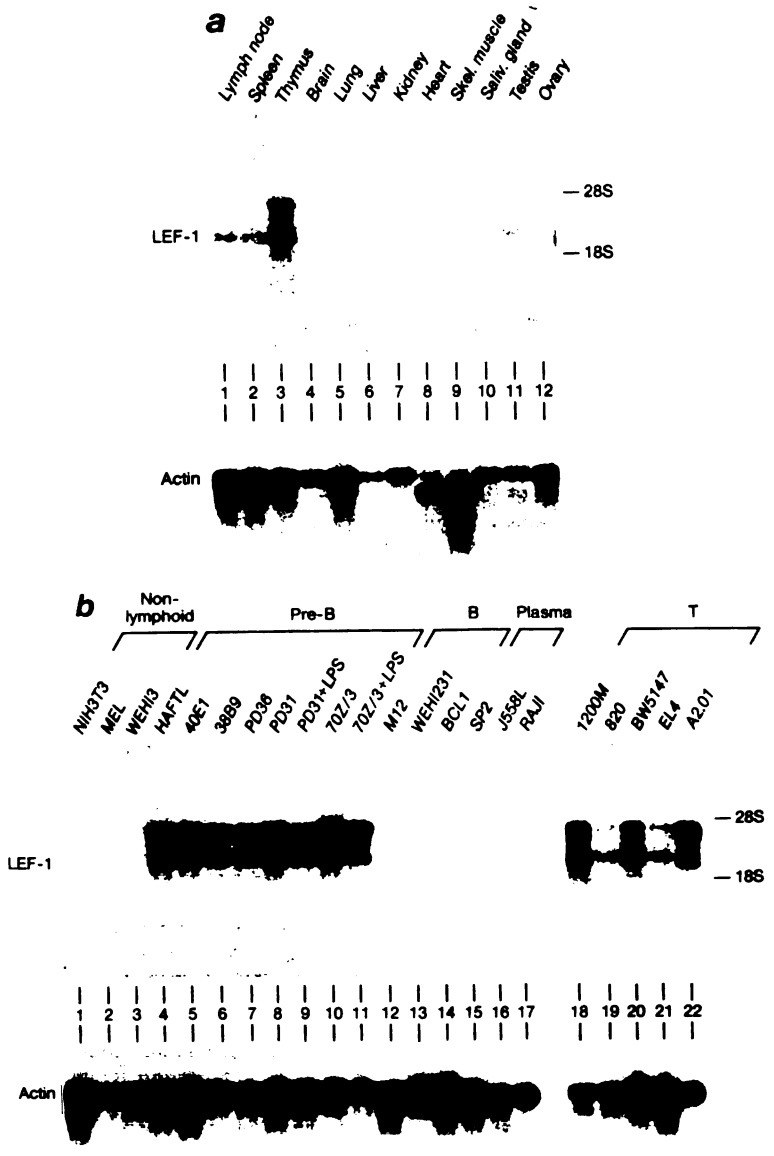


Figure 3. Expression of LEF-1 in mouse tissues and in mouse and human cell lines. (*a*) RNA blot analysis of LEF-1 transcripts in adult mouse tissues. Poly(A)$^+$ RNA (0.5 μg) from various tissues was size-fractionated by gel electrophoresis, transferred to a nylon membrane, and hybridized with a nick-translated DNA probe derived from the GN8 cDNA (nucleotides 1158–1517). The sizes of the major LEF-1 transcripts detected in thymus and at lower abundance in spleen and lymph nodes are ~2.7 and 4.2 kb, as determined by comparison with an 18S and 28S rRNA size marker. The transcripts detected in testis range in size from 3.0 to 3.4 kb. To control for RNA levels, a duplicate RNA blot was hybridized with a nick-translated hamster actin gene probe (*bottom*). (*b*) RNA blot analysis of LEF-1 transcripts in various cell lines. Poly(A)$^+$ RNA (0.5 μg) from various tissue culture cell lines was analyzed for the presence of LEF-1 transcripts as in *a*. RAJI and A2.01 are human lymphoid B- and T-cell lines, respectively. All other cell lines are derived from mouse. They include fetal liver-derived pre-B cells (HAFTL, 40E1, and 38B9), adult bone marrow-derived pre-B cells (PD36, PD31, and 70Z/3), B cells (M12, WEHI231, and BCL1), plasmacytomas (SP2 and J558L), and T cells, including three CD4$^-$/CD8$^-$ cell lines (BW5147, EL4, and A2.01), a CD4$^+$/CD8$^+$ cell line (820), and a CD4$^-$/CD8$^+$ cell line (1200M). Nonlymphoid cell lines include fibroblasts (NIH-3T3), erythroleukemia cells (MEL), and myeloid cells (WEHI 3). The sizes of the LEF-1 transcripts (4.2, 3.7, and 2.7) were estimated by comparison with an 18S and 28S RNA size standard.

gether, these observations suggest that LEF-1 is expressed specifically in pre-B and T lymphocytes and is conserved between mouse and man.

Detection and nuclear localization of endogenous LEF-1

To detect the endogenous LEF-1 protein in lymphoid cells, we raised rabbit polyclonal antibodies to a polypeptide derived from the GL1 cDNA by bacterial overexpression (see Materials and methods). This anti-LEF-1 serum was used to probe immunoblots of whole-cell extracts from various lymphoid and nonlymphoid cells (Fig. 4a, lanes 4–11). A major protein with an apparent molecular mass of 54 kD was detected in pre-B and T cells (lanes 5–8). Minor polypeptides of smaller apparent molecular mass were detected in T cells (lanes 5 and 6), but not in B-lineage cells or in nonlymphoid cells (lanes 4, 7–11). This suggests that the faster-migrating polypeptides represent alternate forms of LEF-1, degradation products, or other T-cell-specific proteins that can cross-react with the anti-LEF-1 serum. For comparison, the anti-LEF-1 serum was reacted with polypeptides that were translated in vitro from GL1- and GN8-derived transcripts (Fig. 4a, lanes 1–3). The GN8-derived polypeptide comigrated with the 54-kD protein detected in pre-B and T cells, supporting the conclusion that the GN8 cDNA contains the entire open reading frame of LEF-1.

The subcellular location of LEF-1 was examined by indirect immunofluorescence microscopy of 70Z/3 pre-B cells reacted with anti-LEF-1 serum (Fig. 4c). Although distinguishing lymphoid cell nuclei and cytoplasm is complicated by their small cytoplasmic volume, nuclear staining could be inferred by examining cells at high magnification (Fig. 4b) and by comparing the immunofluorescence staining of LEF-1 (Fig. 4c) with DAPI staining of DNA (Fig. 4d). LEF-1 staining was nuclear in nondividing cells, but it was distributed throughout the cytoplasm in dividing cells (Fig. 4b–d, cell pair in anaphase at lower right corner and cell in metaphase above it), consistent with their lack of a nuclear envelope. The level of background staining of the cells was determined with preimmune serum (Fig. 4e). These data indicate that LEF-1 is a nuclear protein.

Sequence-specific DNA binding

To determine whether LEF-1 binds DNA in a sequence-specific manner and to identify putative target sequences, we examined various lymphocyte-specific enhancers for LEF-1 binding. Our search was facilitated by two observations. First, activity of the TCRβ enhancer is observed in T and pre-B cells (Krimpenfort et al. 1988;

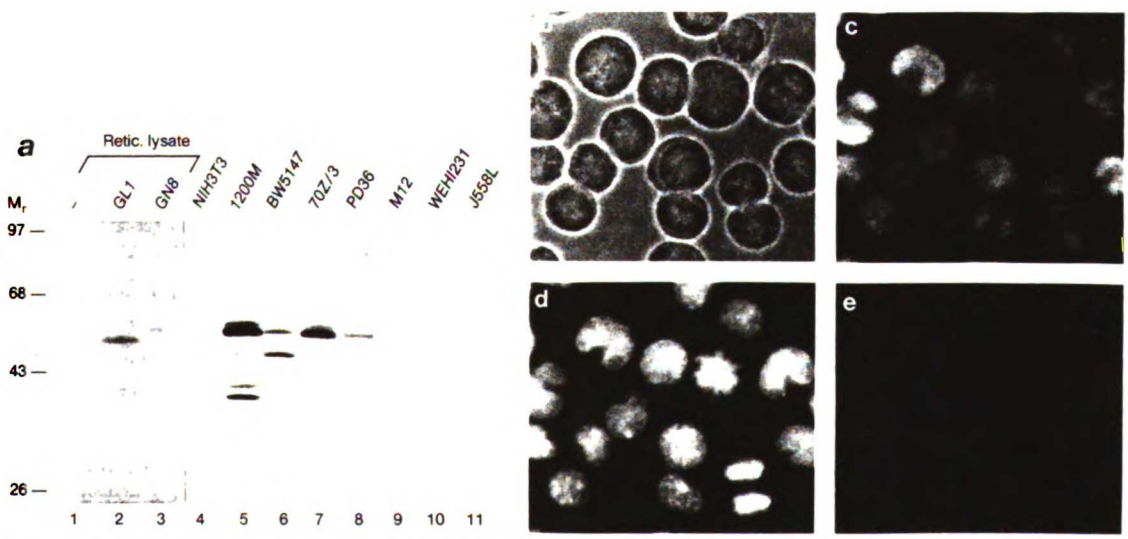


Figure 4. Protein identification and subcellular localization of LEF-1. (*a*) Immunoblot analysis of LEF-1 protein. In vitro-translated recombinant LEF-1 polypeptides (lanes *2* and *3*) and whole-cell protein extracts from various cell lines (lanes *4–11*) were size-fractionated by SDS-PAGE, transferred onto a nylon membrane, and reacted with precleared anti-LEF-1 serum (see Materials and methods) at a dilution of 1 : 3000. Recombinant LEF-1 polypeptides were translated in a reticulocyte lysate from GL1- or GN8-derived transcripts (see Materials and methods). The sizes of molecular mass markers are indicated in kD. (*b–e*) Immunocytochemical staining of endogenous LEF-1. Bright-field (*b*), rhodamine immunofluorescence (*c*), and 4,6-diamidino-2-phenylindole (DAPI) fluorescence (*d*) microscopy of the same field of fixed 70Z/3 pre-B cells reacted with precleared anti-LEF-1 serum at a dilution of 1 : 130 and double-stained with rhodamine-conjugated goat anti-rabbit antibodies and the DNA intercalating agent DAPI. (*e*) Immunofluorescence microscopy of fixed 70Z/3 cells incubated with preimmune serum at a 1 : 130 dilution.

Takeda et al. 1990), thus paralleling the expression pattern of LEF-1. Second, a DNA fragment comprising part of an enhancer associated with the murine CD4 gene can interact with a pre-B- and T-cell-specific nuclear factor in vitro (S. Sawada and D.R. Littman, pers. comm.). Electrophoretic mobility-shift and methylation interference assays performed with in vitro-translated LEF-1 protein indicated that LEF-1 can bind a specific sequence in this CD4 enhancer fragment (A. Amsterdam, S. Sawada, D.R. Littman, and R.Grosschedl, unpubl.). Nucleotide sequences closely related to this LEF-1-binding site were also found to be present in the human and mouse TCRα and TCRβ enhancers (Krimpenfort et al. 1988; McDougall et al. 1988a; Ho and Leiden 1989; Winoto and Baltimore 1989; Gottschalk and Leiden 1990; Takeda et al. 1990). Since the human TCRα enhancer has been well-characterized and delineated to a 116-bp DNA fragment, we selected this enhancer for studies of the binding specificity and the putative regulatory role of LEF-1.

For the studies aimed at examining DNA binding of LEF-1, a recombinant vaccinia virus was generated that allows for overexpression of LEF-1 in infected cells (see Materials and methods). Therefore, infection of HeLa cells, which lack endogenous LEF-1, with this recombinant vaccinia LEF-1 virus or a control vaccinia hemagglutinin (HA) virus enabled us to prepare nuclear extracts that differed only in the presence or absence of LEF-1. Immunoblot analysis of nuclear extracts of HeLa cells infected with the vaccinia LEF-1 virus indicated a level of LEF-1 five times higher than that detected in the T-cell line 1200M (data not shown).

Sequence-specific binding of recombinant LEF-1 to a minimal 98-bp TCRα enhancer fragment (TCRα98; Fig. 5a) was examined by DNase I footprinting with either HeLa/vaccinia LEF-1 or HeLa/vaccinia HA nuclear extracts (Fig 5b, lanes 3 and 4). A DNase I-protected region between nucleotides 62 and 76 was generated only with HeLa/vaccinia LEF-1 nuclear extract (lane 3), indicating sequence-specific binding of recombinant LEF-1 to the TCRα enhancer. Nucleotides flanking the LEF-1-binding site displayed an enhanced sensitivity to DNase I cleavage. Additional protection over a region from nucleotide 21 to 46 was observed with both HeLa cell nuclear extracts (lanes 3 and 4). This footprint coincides with the α_1 footprint that has been observed previously in T-cell nuclear extracts and covers a region containing a consensus cAMP-responsive element (CRE; Jones et al. 1988). The LEF-1-generated footprint coincides with a footprint generated by the recently identified and purified T-cell factor TCF-1α (Waterman and Jones 1990) and overlaps with a larger footprint that has been observed previously in T-cell nuclear extracts and has been termed α_2 (see Fig. 5a; Ho and Leiden 1989).

To gain further insight into LEF-1 binding, the methylation interference pattern of recombinant LEF-1 on a partially methylated 25-bp TCRα DNA fragment was determined using HeLa/vaccinia LEF-1 nuclear extract (Fig. 5a, TCRα25). Binding of LEF-1 is inhibited by methylation of G_{70} and partially inhibited by methylation of G_{69} (Fig. 5c). No methylated G nucleotides on the other strand interfered with LEF-1 binding (data not shown). On the basis of the methylation interference data, TCRα98 and TCRα25 oligonucleotides containing a triple point mutation in the LEF-1-binding site were generated for subsequent studies (Fig. 5a).

Sequence specificity of the LEF-1/TCRα25 protein–DNA interaction was demonstrated by competition of binding with wild-type and mutated TCRα25 oligonucleotides (Fig. 6a). The LEF-1/TCRα25 complex was detected in an electrophoretic mobility-shift assay using HeLa/vaccinia LEF-1 nuclear extract (lane 2) but not with HeLa/vaccinia HA (lane 1). The LEF-1/TCRα25 complex was sensitive to competition with the wild-type oligonucleotide but was resistant to competition with the mutant oligonucleotide (lanes 3–10), demonstrating sequence-specific binding of recombinant LEF-1.

Incubation of the TCRα25 fragment with nuclear extract from the T-cell line 1200M resulted in the formation of a complex that comigrated with a complex containing in vitro-translated LEF-1 (Fig. 6b, lanes 2 and 3). The apparent affinity of the 1200M nuclear protein for the LEF-1-binding site is similar to that of recombinant LEF-1, as determined by competition of binding with wild-type and mutant TCRα25 oligonucleotides (lanes 4–11). To confirm that the complex obtained with 1200M nuclear extract contains LEF-1 or a closely related protein, we incubated the nuclear extract with anti-LEF-1 serum prior to the addition of the labeled TCRα25 oligonucleotide (Fig. 6c). Increasing amounts of serum inhibited the formation of the protein/TCRα25 complex (lanes 4 and 5). No inhibition was observed with preimmune serum (lanes 2 and 3). To control for the specificity of the anti-LEF-1 serum, we examined the effect of anti-LEF-1 serum on the binding of Oct-1 to its OCTA recognition sequence. No interference with binding was observed (lanes 9 and 10). Taken together, these data indicate that both recombinant and endogenous LEF-1 bind specifically to a site in the TCRα enhancer.

Functional importance of LEF-1

The dependence of cell type-specific TCRα enhancer function on the LEF-1-binding site was examined by inserting two copies of the wild-type or a mutated TCRα98 enhancer fragment (see Fig. 5a) 5′ of a tkCAT reporter gene construct containing a 109-bp promoter fragment of the herpes simplex virus (HSV) thymidine kinase (tk) gene linked to the chloramphenicol acetyltransferase (CAT) gene (see Materials and methods). The tkCAT construct contained a triple poly(A$^+$) site upstream of the inserted enhancer fragments to reduce translatable readthrough from vector sequences (Maxwell et al. 1989). Wild-type and mutant TCRα98/tkCAT constructs were transfected together with a Rous sarcoma virus (RSV)–luciferase gene construct as a transfection control into various cell lines (Fig. 7). These experiments indicated that mutation of the LEF-1-binding site decreased enhancer function by a factor of 5 and 10 in the T-cell lines EL4 and BW5147, respectively, and by a factor of 4 in PD36 pre-B cells. In contrast, the mutations did not

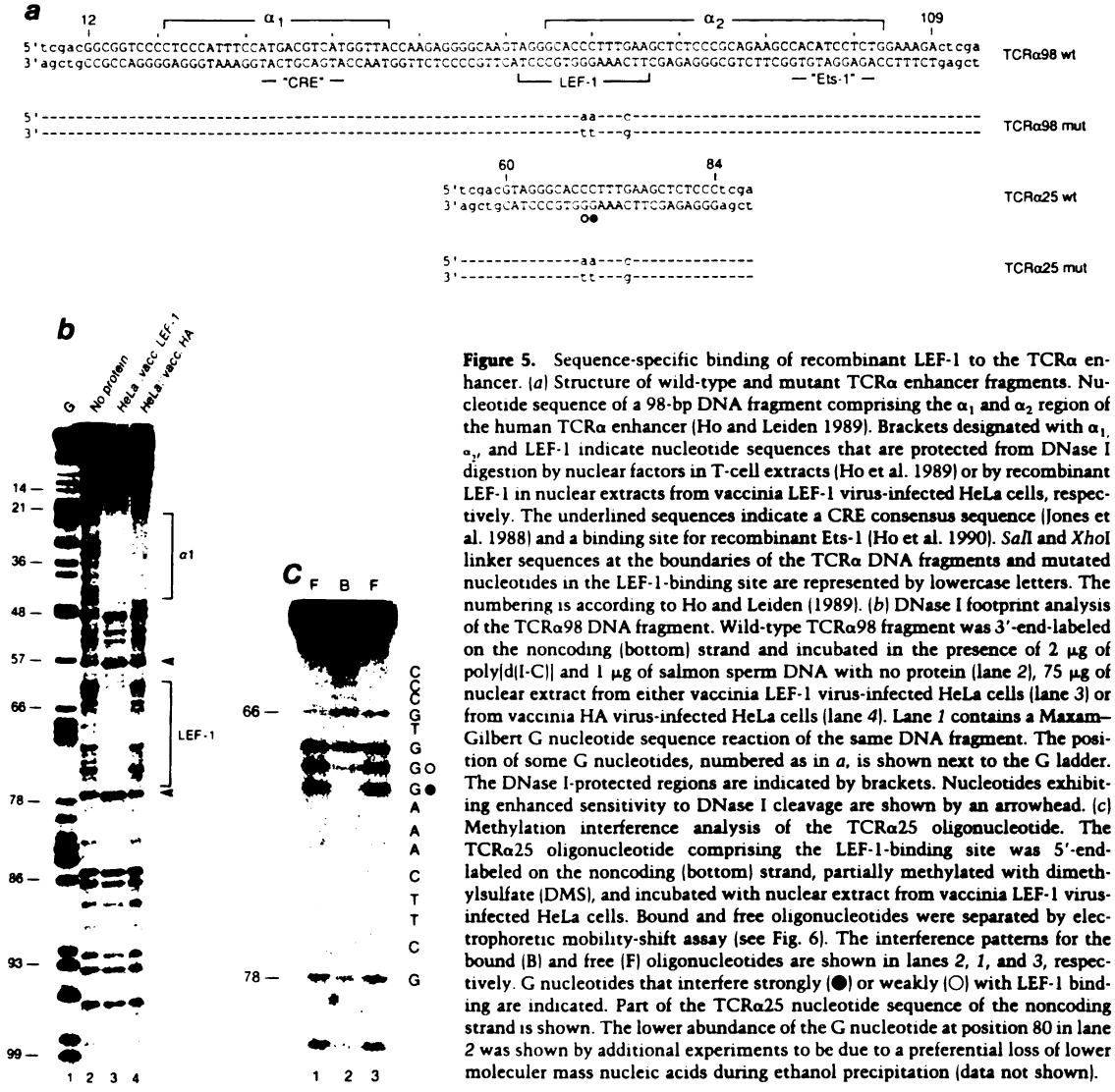


Figure 5. Sequence-specific binding of recombinant LEF-1 to the TCRα enhancer. (*a*) Structure of wild-type and mutant TCRα enhancer fragments. Nucleotide sequence of a 98-bp DNA fragment comprising the α_1 and α_2 region of the human TCRα enhancer (Ho and Leiden 1989). Brackets designated with α_1, α_2, and LEF-1 indicate nucleotide sequences that are protected from DNase I digestion by nuclear factors in T-cell extracts (Ho et al. 1989) or by recombinant LEF-1 in nuclear extracts from vaccinia LEF-1 virus-infected HeLa cells, respectively. The underlined sequences indicate a CRE consensus sequence (Jones et al. 1988) and a binding site for recombinant Ets-1 (Ho et al. 1990). *Sal*I and *Xho*I linker sequences at the boundaries of the TCRα DNA fragments and mutated nucleotides in the LEF-1-binding site are represented by lowercase letters. The numbering is according to Ho and Leiden (1989). (*b*) DNase I footprint analysis of the TCRα98 DNA fragment. Wild-type TCRα98 fragment was 3'-end-labeled on the noncoding (bottom) strand and incubated in the presence of 2 μg of poly[d(I-C)] and 1 μg of salmon sperm DNA with no protein (lane *2*), 75 μg of nuclear extract from either vaccinia LEF-1 virus-infected HeLa cells (lane *3*) or from vaccinia HA virus-infected HeLa cells (lane *4*). Lane *1* contains a Maxam–Gilbert G nucleotide sequence reaction of the same DNA fragment. The position of some G nucleotides, numbered as in *a*, is shown next to the G ladder. The DNase I-protected regions are indicated by brackets. Nucleotides exhibiting enhanced sensitivity to DNase I cleavage are shown by an arrowhead. (*c*) Methylation interference analysis of the TCRα25 oligonucleotide. The TCRα25 oligonucleotide comprising the LEF-1-binding site was 5'-end-labeled on the noncoding (bottom) strand, partially methylated with dimethylsulfate (DMS), and incubated with nuclear extract from vaccinia LEF-1 virus-infected HeLa cells. Bound and free oligonucleotides were separated by electrophoretic mobility-shift assay (see Fig. 6). The interference patterns for the bound (B) and free (F) oligonucleotides are shown in lanes *2*, *1*, and *3*, respectively. G nucleotides that interfere strongly (●) or weakly (○) with LEF-1 binding are indicated. Part of the TCRα25 nucleotide sequence of the noncoding strand is shown. The lower abundance of the G nucleotide at position 80 in lane *2* was shown by additional experiments to be due to a preferential loss of lower molecular mass nucleic acids during ethanol precipitation (data not shown).

affect the low level of CAT expression obtained in M12 B cells, J558 plasmacytomas, and NIH-3T3 fibroblasts. Thus, the maximal TCRα98 enhancer activity depends on the LEF-1-binding site and is observed only in cells that contain LEF-1. The mutation of the LEF-1-binding site in the context of a single copy of the TCRα98 enhancer decreases enhancer function in BW5147 cells by a factor of eight as well, although the overall level of enhancer activity is fivefold lower than that of the duplicated enhancer (data not shown). Interestingly, and in agreement with data concerning the activity of a multimerized TCF-1α-binding site (Waterman and Jones 1990), neither a single nor multimerized TCRα25 oligonucleotide could augment basal transcription from a minimal *fos* promoter (Berkowitz et al. 1989), indicating that the function of the LEF-1-binding site is dependent on the sequence context (data not shown).

The regulatory role of LEF-1 for TCRα enhancer function was examined by cotransfection of M12 B cells, lacking endogenous LEF-1, with the TCRα98/tkCAT reporter gene together with a cytomegalovirus (CMV)/LEF-1 effector plasmid (see Materials and methods). Forced expression of LEF-1 in B cells increased the activity of the wild-type TCRα98 enhancer by a factor of 3.5 relative to the activity of the mutant TCRα98 enhancer carrying a nonfunctional LEF-1-binding site (Fig. 7b). Nei-

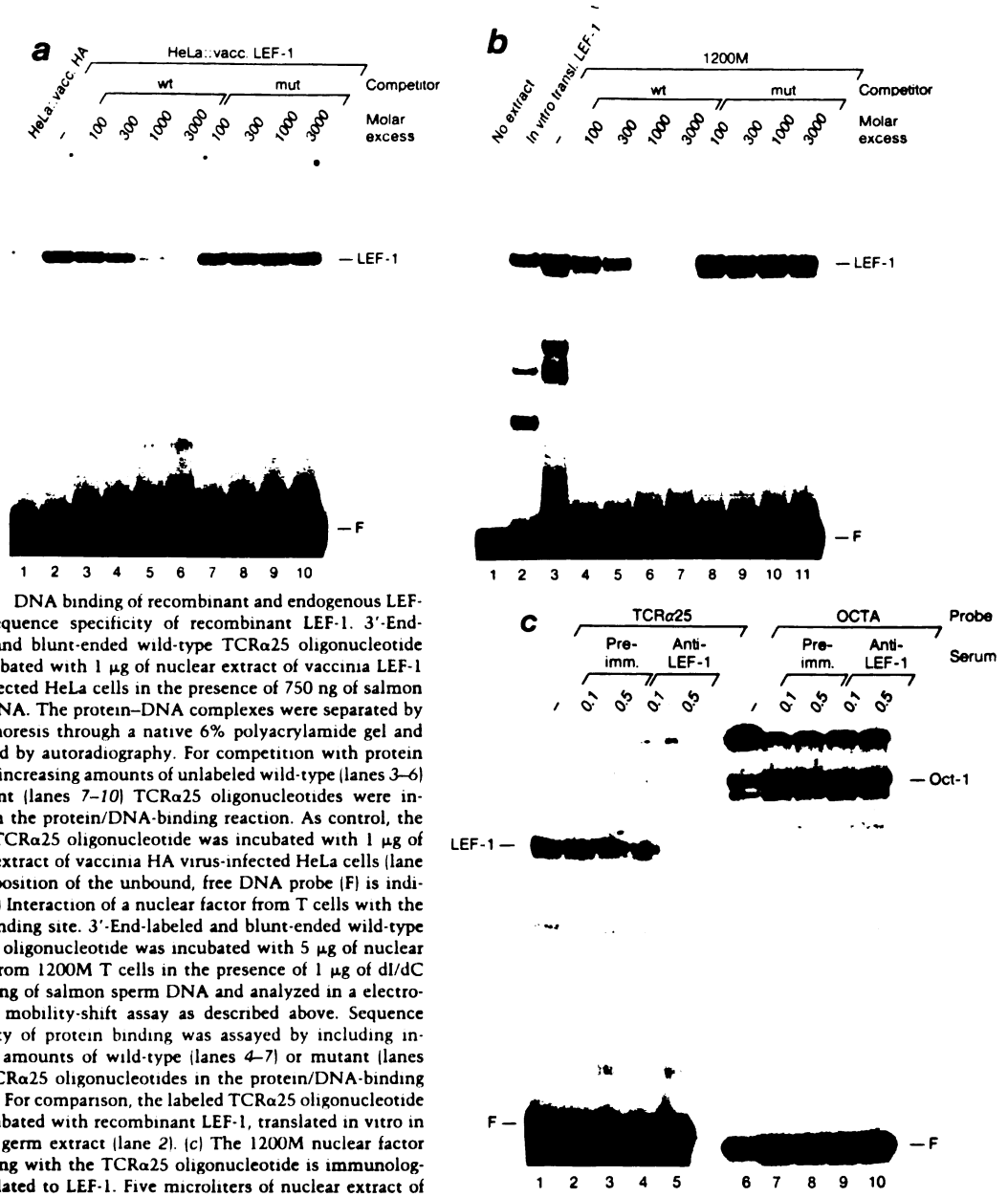


Figure 6. DNA binding of recombinant and endogenous LEF-1. (*a*) Sequence specificity of recombinant LEF-1. 3′-End-labeled and blunt-ended wild-type TCRα25 oligonucleotide was incubated with 1 μg of nuclear extract of vaccinia LEF-1 virus-infected HeLa cells in the presence of 750 ng of salmon sperm DNA. The protein–DNA complexes were separated by electrophoresis through a native 6% polyacrylamide gel and visualized by autoradiography. For competition with protein binding, increasing amounts of unlabeled wild-type (lanes *3–6*) or mutant (lanes *7–10*) TCRα25 oligonucleotides were included in the protein/DNA-binding reaction. As control, the labeled TCRα25 oligonucleotide was incubated with 1 μg of nuclear extract of vaccinia HA virus-infected HeLa cells (lane *1*). The position of the unbound, free DNA probe (F) is indicated. (*b*) Interaction of a nuclear factor from T cells with the LEF-1-binding site. 3′-End-labeled and blunt-ended wild-type TCRα25 oligonucleotide was incubated with 5 μg of nuclear extract from 1200M T cells in the presence of 1 μg of dI/dC and 500 ng of salmon sperm DNA and analyzed in a electrophoretic mobility-shift assay as described above. Sequence specificity of protein binding was assayed by including increasing amounts of wild-type (lanes *4–7*) or mutant (lanes *8–11*) TCRα25 oligonucleotides in the protein/DNA-binding reaction. For comparison, the labeled TCRα25 oligonucleotide was incubated with recombinant LEF-1, translated in vitro in a wheat germ extract (lane *2*). (*c*) The 1200M nuclear factor interacting with the TCRα25 oligonucleotide is immunologically related to LEF-1. Five microliters of nuclear extract of 1200M T cells was incubated with no serum (lane *1*), 0.1 and 0.5 μl of undiluted anti-LEF-1 serum (lanes *4* and *5*), or preimmune serum (lanes *2* and *3*) prior to addition of the 3′-end-labeled and blunt-ended wild-type TCRα25 oligonucleotide. The protein–DNA complexes and unbound, free DNA probe (F) were separated in an electrophoretic mobility-shift assay. One microgram of nuclear extract of 1200M T cells was incubated with no serum (lane 6), preimmune serum (lanes *7* and *8*), or undiluted anti-LEF-1 serum (lanes *9* and *10*), prior to addition of a 5′-end-labeled wild-type OCTA oligonucleotide.

ther the activity of the wild-type nor that of the mutant TCRα98 enhancer was affected by the cotransfected control effector plasmid CMV/inv LEF-1 containing the LEF-1 cDNA insert in an inverse orientation. In contrast, transfection of the CMV/LEF-1 effector plasmid into NIH-3T3 fibroblastic cells did not increase the activity of a cotransfected TCRα98/tkCAT reporter gene (data not shown). Possibly, negative regulation in nonlym-

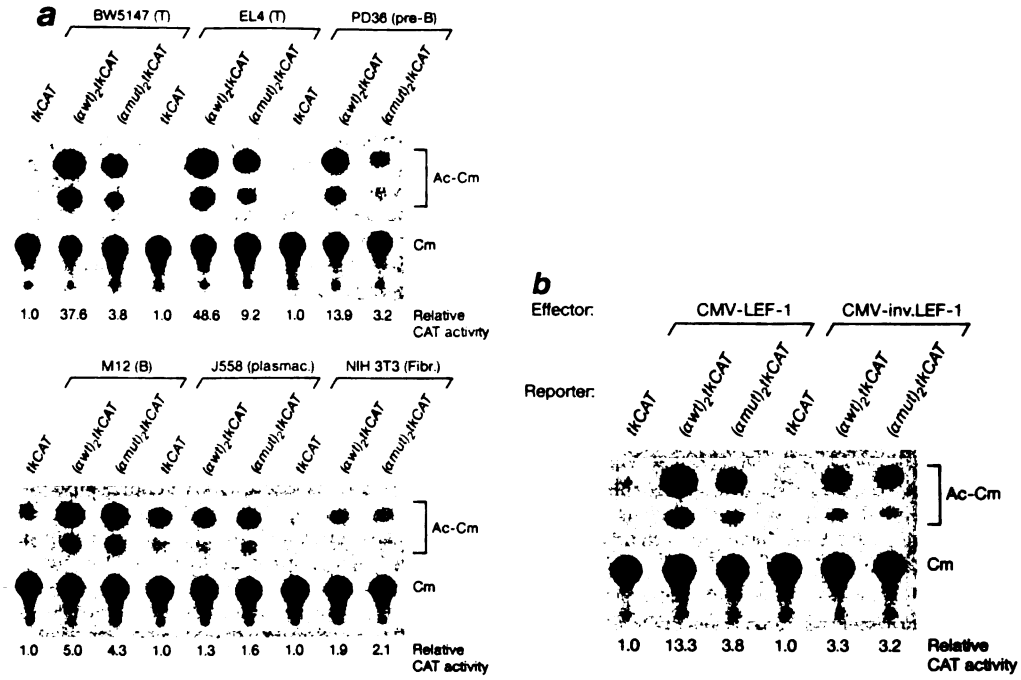


Figure 7. LEF-1 participates in regulating TCRα enhancer activity. (*a*) Dependence of the TCRα98 enhancer on the LEF-1-binding site parallels the expression pattern of LEF-1. ptkCAT and the derivative pTCRα98/tkCAT plasmids, carrying a duplicated wild-type (αwt) or mutant (αmut) TCRα98 enhancer fragment (see Fig. 5a) at position −109 of the HSV tk promoter, were transfected into various cell lines. pRSV–luciferase was included in all transfection experiments as an internal standard. A representative CAT assay, separating the acetyl–chloramphenicol (Ac-Cm) reaction products and chloramphenicol (Cm) by thin-layer chromatography, is shown, and the average relative CAT levels, which were determined by multiple transfections and normalized to the luciferase level, are indicated. The variability between independent experiments was <20%. (*b*) Expression of recombinant LEF-1 augments TCRα enhancer function in a LEF-1-binding site-dependent manner. M12 B-lineage cells, lacking endogenous LEF-1, were cotransfected with the ptkCAT or pTCRα98/tkCAT reporter plasmids together with a CMV/LEF-1 effector plasmid containing the GL1 cDNA linked to the CMV enhancer and promoter. The ratio of reporter–effector plasmid was 1 : 10. The control effector plasmid CMV/inv LEF-1 contains the GL1 cDNA linked to the CMV enhancer–promoter in an inverse orientation. CAT activity was assayed and normalized to the activity of a cotransfected RSV–luciferase gene.

phoid cells prevents LEF-1 from augmenting TCRα98 enhancer function. Alternatively, the function of LEF-1 may be dependent on a lymphoid-specific modification or cooperation with other lymphoid-specific factors. Whatever mechanisms restrict LEF-1 function to lymphoid cells, our experiments indicate that LEF-1 participates in the regulation of cell type-specific function of the TCRα enhancer.

Discussion

In this study we identified a murine regulatory gene that is expressed specifically in pre-B and T lymphocytes. This gene, termed LEF-1, encodes a sequence-specific DNA-binding protein that recognizes a functionally important site in the TCRα enhancer. A regulatory role of LEF-1 in TCRα-enhancer function is based on three arguments. First, transfection of a LEF-1 cDNA expression vector into a mature B-cell line increases the expression of a cotransfected reporter gene, containing a TCRα enhancer, in a LEF-1-binding site-dependent manner. Second, the LEF-1-binding site contributes to the function of this TCRα enhancer. Third, the cell type-specific pattern of maximal enhancer activity parallels the expression of LEF-1. The cell type specificity of both TCRα and TCRβ enhancers, however, is less confined than the expression pattern of their associated genes, suggesting that additional regulation is required to govern the ultimate developmental expression pattern of these genes. Recently, the TCRα gene was shown to contain silencers that appear to restrict gene expression to the $\alpha\beta^+$ subset of T lymphocytes (Winoto and Baltimore 1989b).

The expression of LEF-1 in pre-B and T cells raises the question as to whether LEF-1 is involved in the regulation of other genes that are expressed in early stage B cells and/or T-lineage cells. The binding site of LEF-1 in the human TCRα enhancer coincides with the binding

site of the recently identified and purified nuclear factor TCF-1α (Waterman and Jones 1990). TCF-1α was shown to bind to related sequences associated with the IgH enhancer and the promoters of the *lck*, *CD3γ*, and *CD3δ* genes and human immunodeficiency virus (HIV) (Waterman and Jones 1990). Therefore, we expect LEF-1 to bind to these sites as well. In addition to the coincidence of their recognition site in the TCRα enhancer, the relationship between LEF-1 and TCF-1α is extended further by the similarity of their apparent molecular masses, which are 54 kD for LEF-1 and 53–57 kD for TCF-1α (Waterman and Jones 1990). Moreover, TCF-1α was detected in T-cell lines but not in mature B-cell lines and nonlymphoid cells, which is consistent with the pre-B- and T-cell-specific expression pattern of LEF-1. Finally, a double mutation in the TCRα enhancer that abrogates binding of TCF-1α was recently shown to drastically decrease the function of a duplicated enhancer (Waterman and Jones 1990). Because different proteins can potentially recognize the same nucleotide sequence (Staudt et al. 1986; Scholer et al. 1989; Mitchell and Tjian 1989), further analysis of TCF-1α will be required to determine its relationship to LEF-1.

Although our data are compatible with a putative activator function of LEF-1, neither a single nor multimerized binding sites for LEF-1 augment basal transcription from a minimal *fos* promoter. Instead, function of the LEF-1-binding site can only be detected in the context of other factor-binding sites. This observation suggests either that LEF-1 is dependent on other DNA-binding proteins to bind with high affinity and to exert its function as a transcription factor or that LEF-1 stimulates transcription indirectly by affecting the binding or activity of other transcription factors. A comparison of the formation and stability of LEF-1/DNA complexes generated with nuclear extracts or in vitro-translated LEF-1 and various TCRα enhancer fragments does not indicate a dependence of LEF-1 binding on other factors or factor-binding sites (data not shown). Moreover, the nucleotide sequence context of the LEF-1-binding site is different in the TCRα and TCRβ enhancers. Therefore, we favor the view that LEF-1 may alter the DNA binding or action of other transcription factors at these enhancers.

In addition to its role in the activation of gene expression, LEF-1 may also be involved in negative regulation of gene expression. Recently, a multimerized TCRα enhancer fragment containing the α_2 footprint (see Fig. 5a), which includes the LEF-1-binding site and a binding site for Ets-1 (Ho et al. 1990), was shown to repress gene expression by antagonizing the function of a linked SV40 or RSV enhancer in a T-cell-specific manner (Ho and Leiden 1990b). The molecular components for this repression, however, have not yet been identified, and the putative involvement of LEF-1 has not been examined.

Although our experiments indicate an important role of LEF-1 for cell type-specific TCRα enhancer function, LEF-1 is unlikely to be the only cell type-specific regulatory protein involved. First, mutation of the LEF-1-binding site in the TCRα98 DNA fragment decreases enhancer function only to 10%. The mutant TCRα98 enhancer still displays some degree of cell type specificity because residual enhancer activity can be detected in lymphocytes (excluding plasmacytomas). This observed pattern of enhancer function in lymphocytes parallels the expression of *ets-1* (J. Hagman and R. Grosschedl, unpubl.), which has been shown previously to interact with a site 3′ of the LEF-1-binding site (Ho et al 1990). Second, multiple dispersed point mutations in the $TCR\alpha_2$ footprint, which most likely abrogate binding of both LEF-1 and Ets-1, do not affect the function of a larger TCRα enhancer fragment comprising additional factor-binding sites (Ho and Leiden 1990a). Nevertheless, our data suggest that LEF-1 encodes a pre-B- and T-lymphocyte-specific DNA-binding protein that interacts with a specific site in the TCRα enhancer to participate in the developmental regulation of enhancer function.

The amino acid homology of LEF-1 with the nonhistone HMG-1 protein establishes LEF-1 as a new member of the family of HMG box proteins. In contrast to HMG-1, which is a nonspecific DNA-binding protein (for review, see van Holde 1989), LEF-1 appears to bind DNA in a sequence-specific manner. The relationship between LEF-1 and HMG-1 is reminiscent of that of the prokaryotic proteins IHF and HU which, despite their extensive amino acid homology, bind DNA specifically and nonspecifically, respectively (Yang and Nash 1989). In this respect, LEF-1 also differs from hUBF, another member of the family of HMG box proteins, whose sequence-specific DNA binding to the rRNA promoter was shown to be assisted by the interacting protein SL1 (Bell et al. 1990). We have no evidence that LEF-1 requires an interaction with another protein to bind DNA in a sequence-specific manner. Possibly, the HMG box of LEF-1 evolved into a DNA-binding domain that can recognize a specific nucleotide sequence independent of other proteins. Experiments examining the HMG box of LEF-1 as the putative DNA-binding domain indicated that 94 amino acids comprising the HMG homology are both necessary and sufficient for sequence-specific recognition of the LEF-1-binding site (A. Amsterdam and R. Grosschedl, in prep.).

In conclusion, we have identified a gene that encodes a developmentally regulated sequence-specific DNA-binding protein. In addition to its regulatory role for TCRα enhancer function, we anticipate LEF-1 to participate in the regulation of other genes that specify the lymphocyte phenotype. To unravel the function of LEF-1 for lymphocyte differentiation and mouse development, disruption of this gene in the mouse germ line (Mansour et al. 1988) or expression of dominant-negative mutants in transgenic mice (Herskowitz 1987) will ultimately be required.

Materials and methods

Isolation of cDNA clones

A set of lymphocyte-specific cDNA clones was isolated from a murine 70Z/3 pre-B-cell cDNA library in λgt11 (Ben-Neriah et al. 1986). Lymphocyte-specific clones were identified by screening duplicate bacteriophage lifts with [α-^{32}P]dCTP-labeled first-

strand cDNA generated from either 70Z/3 or MEL murine erythroleukemia poly(A)$^+$ RNA (Bergman et al. 1983; Ben-Neriah et al. 1986). Of 200,000 plaques that hybridized with the radiolabeled 70Z/3 cDNA but failed to hybridize with the radiolabeled MEL cDNA, 10,000 were picked as 16 pools of ~600. These pools were rescreened by the same method using radiolabeled cDNA derived from the murine T-cell lines BW5147 and EL4, the pre-B-cell 70Z/3, and MEL. Single plaques hybridizing with all three labeled lymphoid cDNAs but not with the labeled erythroid cDNA were identified, and the recombinant bacteriophages were isolated (Sambrook et al. 1989). cDNAs were subcloned into Bluescript (Stratagene) for sequencing of both strands of LEF-1 by the dideoxynucleotide method (Sambrook et al. 1989) using synthetic oligonucleotide primers and Sequenase (U.S. Biochemicals).

RNA analysis

Total RNA was prepared from C57BL6 mouse tissues, and total cytoplasmic RNA was prepared from cell lines as described (Bergman et al. 1983; Sambrook et al. 1989). Poly(A)$^+$ RNA was selected by passage over oligo(dT)-cellulose (BMB) (Sambrook et al. 1989). Poly(A)$^+$ RNA samples of 0.5 μg were separated on 1.0% agarose gels containing 2.2 M formaldehyde, transferred to Hybond-N nylon membranes (Amersham), UV-immobilized, and hybridized with a nick-translated 0.36-kb *Hin*cII DNA fragment from the GN8 cDNA (nucleotides 1158–1517 in Fig. 1b). Hybridization was at 42°C in 5× SSC, 20 mM $NaHPO_4$ at pH 6.7, 1× Denhardt's solution, 100 mg/ml of sheared/boiled salmon sperm DNA, and 50% formamide. Washing was in 0.5× SSC and 0.1% SDS at 65°C.

Antibody production, immunoblot analysis, and immunofluorescence

GN8 and GL1 polypeptides were synthesized in a rabbit reticulocyte lysate translation system (Promega) using RNA generated from Bluescript clones of GN8 and GL1 cDNAs with bacteriophage T7 RNA polymerase (BMB). Translation of the T7 transcripts from the GL1 cDNA was initiated at a synthetic in-frame AUG upstream of the cDNA insert, which resulted in an addition of 8 amino acids from the polylinker to the amino terminus of the GL1 polypeptide. Protein samples from cell lines were generated by lysing whole cells in SDS–sample buffer (Harlow and Lane 1988).

Rabbit antibodies were raised against the entire GL1-encoded polypeptide (amino acids 25–397), which was expressed in *Escherichia coli* using the pET vector system (Studier et al. 1990) and purified by preparative SDS–gel electrophoresis and electroelution (Harlow and Lane 1988). The serum was precleared of nonspecific antibodies by diluting it 1 : 10 with 3% BSA in PBS and incubating it with a protein blot of WEHI 231 B-cell extract for 18 hr at 4°C. SDS–gel electrophoresis and immunoblotting was as described (Harlow and Lane 1988), using the precleared antiserum at a final dilution of 1 : 3000. Detection of anti-LEF-1 antibodies was accomplished by incubation of the immunoblot with an alkaline phosphatase-conjugated second antibody (Promega).

For bright-field, rhodamine immunofluorescence, and DAPI fluorescence microscopy, 70Z/3 cells were allowed to settle onto poly-L-lysine-coated slides, fixed in 4% paraformaldehyde for 20 min, and permeabilized with 0.2% Triton X-100 for 4 min, and free formaldehyde groups were reduced in 50 mM NH_4Cl for 4 min (three times) and blocked in 0.5% BSA, 15% goat serum, and 0.05% Tween-20 for 5 min, all in PBS at room temperature. Incubation with the first antibody (anti-GL1 polypeptide) was for 40 min in a humidified chamber. DAPI staining of nuclei was at 0.1 μg/ml. A rhodamine-conjugated goat anti-rabbit antibody (Chemicon) was used for immunodetection.

Recombinant vaccinia viruses

GL1 cDNA encoding amino acids 25–397 of LEF-1 with an added in-frame ATG codon was inserted downstream of the H6 promoter in plasmid HES4, which contains a host-range gene of vaccinia virus (Perkus et al. 1989). Recombinant vaccinia/LEF-1 virus was selected by transfection of the recombinant HES4–GL1 gene construct into CV1 cells and coinfection with the deletion host range mutant vaccinia virus P293 (Perkus et al. 1989). Single plaques were isolated, and recombinant viruses were screened by immunoblot analysis of lysates of infected cells. Vaccinia/HA was generated in the same manner and was generously donated by Drs. George Kemble and Judith White.

Cell culture, nuclear extracts, and viral infections

Cells were grown to a density of 5×10^5 cells/ml. HeLa cells were grown in Joklik modified minimal essential medium (JMEM) supplemented with 10% equine serum. Lymphoid cell lines were grown in RPMI supplemented with 10% fetal calf serum and 50 μM 2-mercaptoethanol.

Nuclear extracts were prepared according to Dingham et al. (1983), with modifications described in Schreiber et al. (1989).

Infection of HeLa cells with recombinant vaccinia virus was performed 48 hr prior to the preparation of the nuclear extracts. For virus infection, 5×10^8 HeLa cells were pelleted and resuspended in 100 ml of JMEM with 2.5% equine serum, and virus was added at a m.o.i. of 1 pfu/cell.

Electrophoretic mobility-shift assays

DNA-binding reactions (10 μl) were carried out as described in the figure legends, containing 5 fmoles of end-labeled TCRα25 DNA probe (labeled on both 3′ ends by Klenow DNA polymerase fill-in reactions with [α-^{32}P]dCTP), poly[d(I-C)] and/or sonicated salmon sperm DNA, 10 mM HEPES (pH 7.9), 50 mM NaCl, 1 mM DTT, and 5% glycerol. Electrophoresis was performed through native 6% polyacrylamide gels in 25 mM Tris, 190 mM glycine, and 5 mM EDTA at 4°C.

DNase I footprint and methylation interference analysis

Nuclear extracts were incubated with 2 fmoles of single end-labeled DNA probe (isolated from a recombinant plasmid and labeled on one 3′ end by Klenow DNA polymerase fill-in reaction with [α-^{32}P]dCTP), for 30 min at room temperature in the presence of 2 μg of poly[d(I-C)], 1 μg of sonicated salmon sperm DNA, 10 mM HEPES (pH 7.9), 50 mM KCl, 5 mM $MgCl_2$, 0.5 mM EDTA, 0.5 mM spermidine, and 10% glycerol. Samples were treated with DNase I at 50 μg/ml for 1 min, and the DNase I digestion was stopped by adjusting the reaction to 25 mM EDTA and 0.2% SDS. DNA samples were phenol-extracted, ethanol-precipitated, and separated by denaturing polyacrylamide gel electrophoresis .

For methylation interference analysis, the TCRα25 oligonucleotide representing the noncoding strand was 5′-end-labeled with [γ-^{32}P]ATP and polynucleotide kinase and annealed with a TCRα25-coding strand, and the double-stranded TCRα25 DNA probe was purified and partially methylated with dimethylsulfate (Maxam and Gilbert 1980). DNA-binding reactions (40 μl) were performed using 50 fmoles of single end-labeled TCRα25 probe, 24 μg of nuclear extract, 8 μg of poly[d(I-C)], 4 μg of sonicated salmon sperm DNA, 10 mM HEPES (pH 7.9), 50 mM

NaCl, 1 mM DTT, and 5% glycerol. Protein–DNA complexes were separated through 6% native polyacrylamide gels and processed as described (Singh et al. 1986).

DNA constructs, DNA transfections, and CAT assays

The tkCAT gene construct was generated by inserting a triple poly(A)$^+$ site fragment from SV40 (Maxwell et al. 1989) into the *Hin*dIII site of pmOTCO containing a 109-bp HSV tk promoter fragment linked to the CAT gene (DeFranco and Yamamoto 1986). To generate the TCRα98 DNA fragments, two 66-bp overlapping oligonucleotides, comprising part of the human TCRα enhancer (nucleotides 12–109; Ho and Leiden 1989) and containing a *Sal*I or *Xho*I linker sequence at their 5′ ends, were annealed, extended, and subcloned into a Bluescript vector. The nucleotide sequence of the cloned wild-type and mutant TCRα98 enhancer fragment (for nucleotide sequence, see Fig. 5a) was confirmed by sequencing, and the TCRα98 DNA fragments were isolated, dimerized, and inserted into the *Sal*I site of ptkCAT. The structure of the pRSV–luciferase gene construct is described in De Wet et al. (1987). For the construction of the CMV/LEF-1 and CMV/invLEF-1 effector plasmids, the isolated and blunt-ended GL1 cDNA insert (encoding amino acids 25–397) was inserted in the sense or antisense orientation into the *Sma*I site of pEV RF2 containing the CMV enhancer/promoter and the translation initiation region from the HSV tk gene (Matthias et al. 1989).

DNA transfections into cell lines were performed as described in Grosschedl and Baltimore (1985), using a DEAE–dextran/chloroquine procedure with 2 μg of DNA/ml (1.8 μg of reporter plus 0.2 μg of pRSV–luciferase, or 0.16 μg of reporter, 1.6 μg of effector, plus 0.2 μg of pRSV–luciferase). After 48 hr, cells were harvested to determine luciferase activity as described (Mangalam et al. 1989). CAT assays were performed as described (Gorman et al. 1982), using 3 μg of protein extract from M12, BW5417, and EL4 cells and 60 μg of protein extract from PD36, J558, and NIH-3T3 cells and incubating it in a CAT cocktail for 2 hr at 37°C. The exact amount of protein extract used was adjusted according to the luciferase activity of each sample. Acetylated [^{14}C]chloramphenicol was separated by thin-layer chromatography and autoradiographed. For quantitation of the acetylated [^{14}C]chloramphenicol, the chromatogram was exposed on a PhosphorImager (Molecular Dynamics), allowing direct determination of the radioactivity in each spot.

Acknowledgments

We are grateful to Y. Ben-Neriah for providing the 70Z/3 cDNA library, G. Kemble and J. White for their generous gift of vaccinia–HA virus, and E.R. Richie and K. Yamamoto for providing the 1200M, and 820 cell lines and the mOTCO plasmid, respectively. We thank I. Herskowitz for discussions, S. Sawada and D. Littman for sharing their unpublished data concerning the binding of a pre-B and T-cell-specific factor to the CD4 enhancer, C. Staben for communicating the amino acid sequence of mt a1 prior to publication, and D. Ornelles for advice on immunofluorescence microscopy. We also thank Drs. Nancy Craig, William Forrester, Ira Herskowitz, Dan Littman, and Keith Yamamoto for thoughtful comments on the manuscript. C.B. is supported by a postdoctoral fellowship from Fonds pour la formation du Chercheurs et l'Aide a la Recherche (FCAR). This work was supported by funds from the Howard Hughes Medical Institute to R.G.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

Note added in proof

Since submission of the manuscript, the cloning of cDNAs encoding a human T-lymphocyte-specific HMG-box protein, termed TCF-1, was reported by Van de Wetering (*EMBO J.* **10:** 123–132 (1991)). The central 78 amino acids of the HMG box of LEF-1 are 97% identical with those of TCF-1. However, the proteins appear to be encoded by distinct genes. In particular, the amino acid sequences of the proteins excluding the central region of the HMG box are different. Moreover, the nucleotide sequences of the human TCF-1 and murine LEF-1 in the actual region of the HMG box are divergent in the third nucleotide positions of most codons. Finally, the RNA blot analysis indicate a distinct pattern for TCF-1 and LEF-1 mRNAs. We also learned of the cloning of cDNAs encoding the human T-cell factor TCF-1α by Waterman et al. (*Genes & Dev.* **5:** 656–669), which appears to be the human homolog of LEF-1. The nucleotide sequence of murine LEF-1 has been submitted to the EMBL/GenBank data base libraries.

The originally submitted nucleotide sequence of LEF-1 contained an erroneous 13-necleotide insertion at position 2143 due to a subcloning artifact. The nucleotide sequence and the deduced amino acid sequence of LEF-1 shown in Figure 1 represent the corrected version.

References

Banerji, T., L. Olson, and W. Schaffner. 1983. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell* **33:** 729–740.

Bell, S.P., H.-M. Jantzen, and R. Tjian. 1990. Assembly of alternative multiprotein complexes directs RNA promoter selectively. *Genes & Dev.* **4:** 943–954.

Ben-Neriah, Y., A. Bernards, M. Paskind, G.Q. Daley, and D. Baltimore. 1986. Alternative 5′ exons in c-*abl* mRNA. *Cell* **44:** 577–586.

Bergman, Y., S.J. Stewart, S. Levy, and R. Levy. 1983. Biosynthesis, glycosylation, and in vitro translation of the human T cell antigen lev-4. *J. Immunol.* **131:** 1876–1881.

Berkowitz, L.A., K.T. Riabowol, and M.Z. Gilman. 1989. Multiple sequence elements of a single functional class are required for cyclic AMP responsiveness of the mouse c-*fos* promoter. *Mol. Cell. Biol.* **9:** 4272–4281.

Blackman, M., J. Kappler, and P. Marrack. 1990. The role of the T cell receptor in positive and negative selection of developing T cells. *Science* **248:** 1335–1341.

Blackwell, T.K. and F.W. Alt. 1988. Immunoglobulin genes. In *Molecular immunology* (ed. B.D. Hames and D.M. Glover), pp. 1–60. IRL Press, Washington, D.C.

Cooper, M.D. and P.D. Burrows. 1989. B-cell differentiation. In *Immunoglobulin genes* (ed T. Honjo, F.W. Alt, and T.H. Rabbits), pp.1–22. Academic Press, London.

Davis, M.M. and P.J. Bjorkman. 1988. T-cell antigen receptor genes and T-cell recognition. *Nature* **334:** 395–401.

DeFranco, D. and K.R. Yamamoto. 1986. Two different factors act separately or together to specify functionally distinct activities at a single transcriptional enhancer. *Mol. Cell. Biol.* **6:** 993–1001.

De Wet, J.R., K.V. Wood, M. DeLuca, D.R. Helinski, and S. Subramani. 1987. Firefly luciferase gene: Structure and expression in mammalian cells. *Mol. Cell. Biol.* **7:** 725–737.

Dingham, J.D., R.M. Lebovitz, and R.G. Roeder. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11:** 1475–1489.

Gerster, T., D. Picard, and W. Schaffer. 1986. During B-cell dif-

ferentiation enhancer activity and transcription rate of immunoglobulin heavy chain genes are high before mRNA accumulation. *Cell* **45:** 45–52.

Gillies, S.D., S.L. Morrison, V.T. Oi, and S. Tonegawa. 1983. A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell* **33:** 717–728.

Gorman, C.M., L.F. Moffat, and B.H. Howard. 1982. Recombinant genomes which express chloramphenicolacetyl transferase in mammalian cells. *Mol. Cell. Biol.* **2:** 1044–1051.

Gottschalk, L. and J.M. Leiden. 1990. Identification and functional characterization of the human T-cell receptor β gene transcriptional enhancer: Common nuclear proteins interact with the transcriptional regulatory elements of the T-cell receptor α and β genes. *Mol. Cell. Biol.* **10:** 5486–5495.

Grosschedl, R. and D. Baltimore. 1985. Cell-type specificity of immunoglobulin gene expression is regulated by at least three DNA sequence elements. *Cell* **41:** 885–897.

Gubbay, J., J. Collignon, P. Koopman, B. Capel, A. Economou, A. Münsterberg, N. Vivian, P. Goodfellow, and R. Lovell-Badge. 1990. A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature* **346:** 245–250.

Harlow, E. and D. Lane. 1988. *Antibodies: A laboratory manual* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Henthorn, P., M. Kiledjian, and T. Kadesch. 1990. Two distinct transcription factors that bind the immunoglobulin enhancer μE5/κE2 motif. *Science* **247:** 467–470.

Herskowitz, I. 1987. Functional inactivation of genes by dominant negative mutations. *Nature* **329:** 219–222.

Ho, I.-C. and J.M. Leiden. 1990a. Regulation of the human T cell receptor α gene enhancer: Multiple ubiquitous and T cell-specific nuclear proteins interact with four hypomethylated enhancer elements. *Mol. Cell. Biol.* **10:** 4720–4727.

Ho, I.-C. and J.M. Leiden. 1990b. The Tα2 nuclear protein binding site from the human T cell receptor α enhancer functions as both a T cell-specific transcriptional activator and repressor. *J. Exp. Med.* **172:** 1443–1449.

Ho, I.-C., L.-H. Yang, G. Morle, and J.M. Leiden. 1989. A T-cell-specific transcriptional enhancer element 3′ of the Cα in the human T-cell receptor α locus. *Proc. Natl. Acad. Sci.* **86:** 6714–6718.

Ho, I.-C., N.K. Bhat, L.R. Gottschalk, T. Lindsten, C.B. Thompson, T.S. Papas, and J.M. Leiden. 1990. Sequence-specific binding of human Ets-1 to the T cell receptor α gene enhancer. *Science* **250:** 814–818.

Jantzen, H.-M., A. Admon, P.B. Bell, and R. Tjian. 1990. Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins. *Nature* **344:** 830–836.

Jones, N.C., P.W.J. Rigby, and E.B. Ziff. 1988. *Trans*-acting protein factors and the regulation of eukaryotic transcription: Lessons from studies on DNA tumor viruses. *Genes & Dev.* **2:** 267–281.

Kelly, M., J. Burke, M. Smith, A. Klar, and D. Beach. 1988. Four mating-type genes control sexual differentiation in the fission yeast. *EMBO J.* **7:** 1537–1547.

Krimpenfort, P., R. de Jong, Y. Uematsu, Z. Dembic, S. Ryser, H. von Boehmer, M. Steinmetz, and A. Berns. 1988. Transcription of T cell receptor β-chain genes is controlled by a downstream regulatory element. *EMBO J.* **7:** 745–750.

Mangalam, H.J., V.R. Albert, H.A. Ingraham, M. Kapiloff, L. Wilson, C. Nelson, H. Elsholtz, and M.G. Rosenfeld. 1989. A pituitary POU domain protein, Pit-1, activates both growth hormone and prolactin promoters transcriptionally. *Genes & Dev.* **3:** 946–958.

Mansour, S.L., K.R. Thomas, and M.R. Capecchi. 1988. Disruption of the proto-oncogene *int-2* in mouse embryo-derived stem cells: A general strategy for targeting mutations to non-selectable genes. *Nature* **336:** 348–352.

Matthias, P., M.M. Muller, E. Schreiber, S. Rusconi, and W. Schaffner. 1989. Eukaryotic expression vectors for the analysis of mutant proteins. *Nucleic Acids Res.* **17:** 6418.

Maxam, A.M. and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65:** 499–550.

Maxwell, I.H., G.S. Harrison, W.M. Wood, and F. Maxwell. 1989. A DNA cassette containing a trimerized SV40 polyadenylation signal which efficiently blocks spurious plasmid-initiated transcription. *BioTechniques* **7:** 276–280.

McDougall, S., C.L. Peterson, and K. Calame. 1988. A transcriptional enhancer 3′ of the Cβ2 in the T cell receptor β locus. *Science* **241:** 205–208.

Mitchell, P.J. and R. Tjian. 1989. Transcriptional regulation in mammalian cells by sequence specific DNA binding proteins. *Science* **245:** 371–378.

Muller, M.M., S. Ruppsert, W. Schaffner, and P. Matthias. 1988. A cloned octamer transcription factor stimulates transcription from lymphoid-specific promoters in non-B cells. *Nature* **336:** 544–551.

Murre, C., P. Schonleber McCaw, and D. Baltimore. 1989. A new DNA-binding and dimerization motif in immunoglobulin enhancer binding, daughterless, myoD, and myc proteins. *Cell* **56:** 777–783.

Nelson, K., E. Mather, and R. Perry. 1984. Lipopolysaccharide-induced transcription of the κ immunoglobulin locus occurs on both alleles and is independent of methylation status. *Nucleic Acids Res.* **12:** 1911–1923.

Perkus, M.E., K. Limbach, and E. Paoletti. 1989. Cloning and expression of foreign genes in vaccinia virus, using a host range selection system. *J. Virol.* **63:** 3829–3836.

Roman, C., J.S. Platero, J. Shuman, and K. Calame. 1990. Ig/EBP-1: A ubiquitously expressed immunoglobulin enhancer binding protein that is similar to C/EBP and heterodimerizes with C/EBP. *Genes & Dev.* **4:** 1404–1415.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: A laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Scheidereit, C., J.A. Cromlish, T. Gerster, K. Kawakami, C. Balmaceda, R.A. Currie, and R.G. Roeder. 1988. A human lymphoid-specific transcription factor that activates immunoglobulin genes is a homeo-box protein. *Nature* **336:** 551–557.

Scholer, H.R., A.K. Hatzopoulos, R. Balling, N. Suzuki, and P. Gruss. 1989. A family of octamer-specific proteins present during mouse embryogenesis: Evidence for germline-specific expression of an Oct factor. *EMBO J.* **8:** 2543–2550.

Schreiber, E., P. Matthias, M.M. Muller, and W. Schaffner. 1989. Rapid detection of octamer binding proteins with "mini-extracts" prepared from a small number of cells. *Nucleic Acids Res.* **17:** 6419.

Sinclair, A.H., P. Berta, M.S. Palmer, J.R. Hawkins, B.L. Griffiths, M.J. Smith, J.W. Foster, A.M. Frischauf, R. Lovell-Badge and P.N. Goodfellow. 1990. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* **346:** 240–244.

Singh, H., J.H. LeBowitz, A.S. Baldwin, and P.A. Sharp. 1986. A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. *Nature* **319:** 154–158.

Staben, C. and C. Yanofsky. 1990. *Neurospora crassa a* mating-type region. *Proc. Natl. Acad. Sci.* **87:** 4917–4921.

Staudt, L.M., H. Singh, R. Sen, T. Wirth, P.A. Sharp, and D.

Baltimore. 1986. A lymphoid-specific protein binding to the octamer motif of immunoglobulin genes. *Nature* **323:** 640–643.

Staudt, L.M., R.G. Clerc, J.H. LeBowitz, P.A. Sharp, and D. Baltimore. 1988. Cloning of a lymphoid-specific cDNA encoding a protein binding the regulatory octamer DNA motif. *Science* **241:** 577–580.

Studier, F.W., A.H. Rosenberg, J.J. Dunn, and J.W. Dubendorff. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185:** 60–89.

Takeda, J., A. Cheng, F. Mauxion, C.A. Nelson, R.D. Newberry, W.C. Sha, R. Sen, and D. Loh. 1990. Functional analysis of the murine T cell receptor β enhancer and characteristics of its DNA-binding proteins. *Mol. Cell. Biol.* **10:** 5027–5035.

van Holde, K.E. 1989. Chromatin (ed. A. Rich). Springer-Verlag Press, New York.

Waterman, M.L. and K.A. Jones. 1990. Purification of TCR-1α, a T-cell-specific transcription factor that activates the T-cell receptor Cα gene enhancer in a context-dependent manner. *New Biol.* **2:** 621–636.

Wen, L., J.-K. Huang, B.H. Johnson, and G.R. Reeck. 1989. A human placental cDNA clone that encodes non-histone chromosomal protein HMG-1. *Nucleic Acids Res.* **17:** 1197–1214.

Winoto, A. and D. Baltimore. 1989a. A novel, inducible and T cell-specific enhancer located at the 3′ end of the T cell receptor α locus. *EMBO J.* **8:** 729–733.

———. 1989b. αβ lineage-specific expression of the α T cell receptor gene by nearby silencers. *Cell* **59:** 649–655.

Yang, C.C. and H. Nash. 1989. The interaction of E. coli IHF protein with its specific binding sites. *Cell* **57:** 869–880.

CHAPTER THREE:

Heterogeneously initiated transcription from the Pre-B- and B-cell-specific *mb-1* promoter: analysis of the requirement for upstream factor-binding sites and initiation site sequences.

MOLECULAR AND CELLULAR BIOLOGY, Nov. 1991, p. 5756–5766
0270-7306/91/115756-11$02.00/0

Vol. 11, No. 11

Heterogeneously Initiated Transcription from the Pre-B- and B-Cell-Specific *mb-1* Promoter: Analysis of the Requirement for Upstream Factor-Binding Sites and Initiation Site Sequences

ADAM TRAVIS, JAMES HAGMAN, AND RUDOLF GROSSCHEDL*

Departments of Microbiology and Immunology and of Biochemistry and Biophysics, Howard Hughes Medical Institute, University of California, San Francisco, California 94143-0414

Received 24 June 1991/Accepted 20 August 1991

The *mb-1* gene, encoding a membrane immunoglobulin-associated protein, is developmentally regulated and expressed specifically in pre-B and mature B lymphocytes. Analysis of the TATA-less *mb-1* promoter indicated that it directs initiation of transcription from multiple sites. Promoter sequences between −68 and +70 conferred the correct pattern of cell type-specific transcription upon a heterologous gene. Two nuclear factor-binding sites that are important for promoter function were identified between −59 and −38. Both sites interacted with ubiquitous nuclear factors in vitro. One of these factors was identified as Sp1. Multimerized copies of both factor-binding sites augmented expression from a heterologous minimal promoter in both lymphoid and nonlymphoid cells, suggesting that additional *mb-1* promoter sequences are involved in determining the correct cell type specificity. Analysis of the heterogeneity of transcription initiation indicated that a mutation which increased the distance between upstream sequences and the region of initiation resulted in the utilization of a novel set of initiation sites. Moreover, an insertion of a TATA element into the *mb-1* promoter at −30 biased initiation of transcription to +1 but did not abolish the use of the other sites. Mutation of an initiator sequence homology encompassing one of the major initiation sites had only a minor effect on its utilization. From these data, we conclude that upstream factor-binding sites in the TATA-less *mb-1* promoter define a region in which initiation of transcription occurs at multiple sites.

B-lymphocyte differentiation turns multipotential progenitor cells into highly specialized cells which ultimately secrete antibody (reviewed in references 1 and 13). The B-cell lineage can be divided into three major stages of differentiation which have been largely defined by changes in the pattern of immunoglobulin (Ig) gene expression. Pre-B cells represent the earliest stage in which the Ig μ heavy-chain gene is expressed. Subsequently, the Ig κ or λ light-chain gene is expressed in B cells, which display antibody on the cell surface. Together, pre-B and B cells represent the early, antigen-independent stages of B-cell differentiation. After B cells encounter antigen, they terminally differentiate into antibody-secreting plasma cells.

Antibody production requires the expression of other lymphoid-specific genes that can be classified according to their stage-specific expression pattern. One set of genes is expressed in pre-B cells but not in later-stage B cells. Included in this group are the λ_5 and V_{preB} genes (28, 42), which likely encode the Ig ω and ι surrogate light-chain components (9), as well as the terminal deoxynucleotidyl transferase (TdT) gene (31) and the recombinase-activating genes (*RAG1* and *RAG2*) (39, 46) which are involved in the assembly of Ig gene segments. The other set of genes is expressed specifically in pre-B and surface IgM-positive B cells, but not in late-stage antibody-secreting plasma cells, and includes the *mb-1* and CD19 genes (43, 52). The *mb-1* gene encodes a protein which is noncovalently associated with membrane-bound IgM (6, 24) and may function as part of a complex involved in signal transduction (5). Likewise, CD19 is a membrane protein thought to participate in signal transduction (11).

Most studies of gene regulation during B-cell differentiation have focused on the lineage-specific expression of the Ig heavy- and light-chain genes. The cell type-specific function of the promoters of Ig genes was shown to be governed by the OCTA site, a nucleotide sequence which can interact with the lymphoid-specific Oct-2 protein and the ubiquitous Oct-1 protein (36, 47, 48, 51, 55). In addition to the OCTA site, the promoters of the Ig genes contain a TATA element positioned approximately 30 bp upstream of the transcription start site. TATA elements are recognized by the general transcription factor TFIID (14, 38, 45), and they have been shown to specify the site of transcription initiation (2, 20, 21) and to be critical for promoter activity in vitro (12, 33, 54).

To address the molecular mechanisms that underlie the common lymphoid lineage specificity and distinct temporal expression patterns of the genes involved in antibody production, we began to study the cell type-specific regulation of the *mb-1* gene. The promoter of this gene differs from those of Ig genes in two aspects. First, the *mb-1* promoter lacks an OCTA site, raising the question as to whether other cell type-specific factors are involved in its regulation. Second, the *mb-1* promoter is a representative of a group of promoters of lymphoid-specific genes that lack a discernible TATA element (25). Other members of this group include the B-cell-specific genes λ_5 (29), V_{preB} (28), and B29 (23) and the early B- and T-cell-specific TdT gene (49). Similar to other TATA-less promoters, transcription from the promoters of the *mb-1*, B29, and V_{preB} genes initiates at multiple sites (this study; 23, 52a). By contrast, transcription from the TATA-less TdT promoter initiates at a single nucleotide within an initiator sequence, which appears to substitute for a TATA element (49, 50). Therefore, the question arises as to whether heterogeneous initiation from the *mb-1* promoter

* Corresponding author.

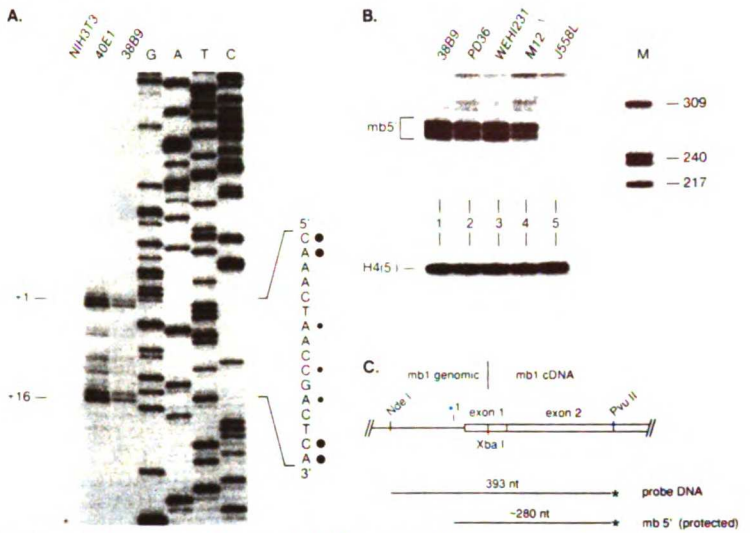


FIG. 1. Evidence that transcription of the *mb-1* gene is initiated at multiple sites. (A) Primer extension mapping of the 5′ ends of endogenous *mb-1* transcripts in RNA from two pre-B-cell lines (40E1 and 38B9) and a fibroblast cell line (NIH 3T3). A DNA sequencing ladder (G, A, T, C) and the sequence of the coding strand at the sites of initiation are shown. Large and small filled circles designate strong and weak initiation sites, respectively. The 5′-most major primer-extended product was arbitrarily designated +1. (B) Detection of the 5′ ends of endogenous *mb-1* transcripts (mb5′) by S1 nuclease protection analysis. RNA from pre-B-cell lines (38B9 and PD36), B-cell lines (WEHI231 and M12), and a plasmacytoma cell line (J558L) was analyzed. Below, an S1 nuclease protection assay detecting the 5′ ends of endogenous H4 transcripts [H4(5′)] is shown. (C) Schematic diagram of the 5′-end-labelled DNA probe used for detection of the *mb-1* transcripts by S1 nuclease mapping. The DNA probe was isolated from a gene construct containing a fusion of genomic and cDNA sequences in exon 1 (see Materials and Methods). nt, nucleotides.

differs mechanistically from that mediated by the TdT initiator, which specifies a single start site.

Here, we report that the *mb-1* promoter region between −68 and +70 directs early B-cell-specific transcription from multiple sites. Transcription from the *mb-1* promoter is dependent on two upstream factor binding sites. Both sites interact with ubiquitous nuclear factors in vitro and can augment the activity of a heterologous minimal promoter in lymphoid and nonlymphoid cells. Heterogeneous initiation of transcription from the *mb-1* promoter was found to depend on the sequence of the initiation sites and their approximate location relative to upstream sequences.

MATERIALS AND METHODS

Isolation of the *mb-1* promoter. Genomic clones of the *mb-1* gene were isolated from a BALB/c mouse genomic library in the bacteriophage λ vector EMBL-3 SP6/T7 (Clontech), using a ^{32}P-labelled *mb-1* cDNA clone as the probe. Upon digestion of the *mb-1* genomic bacteriophage clone M9, a 4.5-kb *Bam*HI-*Xho*I fragment containing the *mb-1* promoter region was identified by using a ^{32}P-labelled synthetic oligonucleotide comprising the noncoding strand of the *mb-1* cDNA from positions +63 to +43 (coordinates of cDNA as in reference 43): 5′-ACAGGCGTATGACAAGA AGAG. This fragment was subcloned into Bluescript M13+ (Stratagene) to yield plasmid pM9-4. The nucleotide sequence of the *mb-1* promoter was determined by using this oligonucleotide, plasmid pM9-4, Sequenase (U.S. Biochemical), and published methods (44), and it was found to match the sequence previously obtained (25) at 318 of 323 nucleotide positions (Fig. 3A). Sequences 5′ to the *Bam*HI site (position −252; Fig. 3A) were isolated from the M9 *mb-1* genomic bacteriophage clone as a 2.5-kb *Bam*HI fragment for use in the construction of −800βPy and −2800βPy (see below).

Plasmid construction and mutagenesis. To generate the βPy vector backbone (Fig. 2A), the *Nco*I site at +359 of a hybrid mouse-human β-globin gene (7) was blunt ended with Klenow enzyme, *Sal*I linkers were ligated to the blunt ends, and a 1.8-kb *Sal*I-*Xho*I fragment containing the hybrid β-globin gene sequences from +359 to +2200 (37) was isolated and ligated with the *Sal*I-*Xho*I vector fragment of plasmid pμ (19).

Isolation of the *mb-1* promoter region for expression analysis was accomplished by first introducing a *Hin*dIII site into the first exon (25) of the *mb-1* gene at position +70 (Fig. 3A) by site-directed mutagenesis. Single-stranded DNA was prepared from pM9-4, and mutagenesis was performed as described previously (30), using the oligonucleotide 5′-CCTGC CTCTCCTCCTGCAAGCTTCATACGCCTGTTTG, yielding plasmid pM9-4S1. The 325-bp *Bam*HI-*Hin*dIII fragment (Fig. 3A) was then isolated from pM9-4S1, blunt ended with Klenow enzyme, and ligated into blunt-ended *Sal*I βPy vector to yield wt(−252)βPy [abbreviated wt(−252)]. The constructs Δ−218, Δ−159, Δ−114, Δ−68, and Δ−22 were similarly generated by subcloning blunt-ended *Hph*I-*Hin*dIII, *Bal*I-*Hin*dIII, *Nde*I-*Hin*dIII, *Hpa*II-*Hin*dIII, and *Asp* 718-*Hin*dIII fragments, respectively, into the blunt-ended *Sal*I βPy vector. An identical set of constructs lacking polyomavirus sequences was generated for transfection of M12 B cells (Fig. 3C) by subcloning the same set of frag-

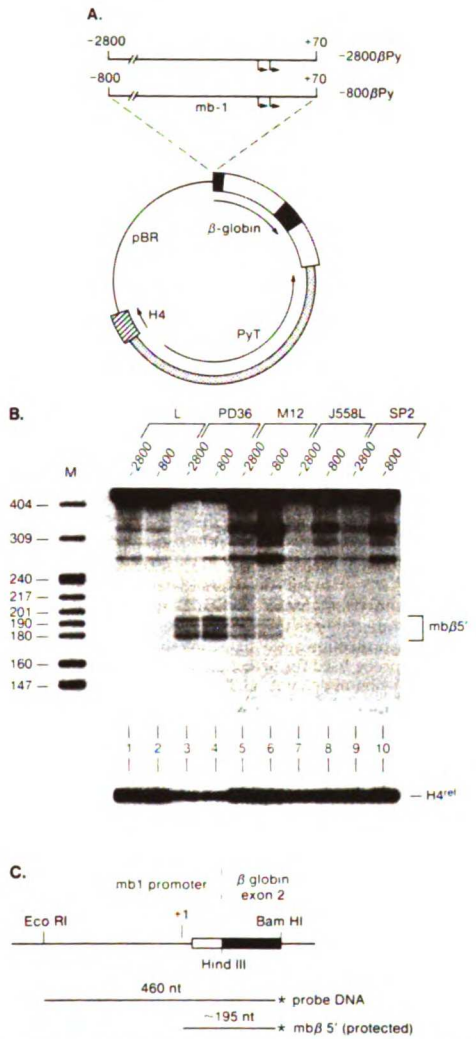


FIG. 2. Lineage- and differentiation stage-specific activity of the *mb-1* gene promoter. (A) Schematic diagram of the plasmid backbone of the *mb-1* promoter constructs used in this and subsequent transfection experiments. The constructs contain *mb-1* promoter fragments linked to the second exon of β-globin genomic sequences. The second and third exons of the β-globin gene are represented by filled bars. The gene constructs also contain the early region of the polyomavirus (PyT) and an H4 reference gene (H4) (19). (B) Detection of transcripts from the *mb-1* promoter constructs by S1 nuclease protection analysis. Constructs containing *mb-1* promoter sequences from −2800 to +70 bp or from −800 to +70 bp (shown in panel A) were transfected into fibroblasts (L), pre-B cells (PD36), B cells (M12), and plasmacytoma cells (J558L and SP2). Accurately initiated transcripts are indicated (mbβ5'). Higher-molecular-weight bands represent read-through transcripts from a strong upstream promoter in the vector. Transcription from this upstream promoter, however, does not interfere with the activity of the *mb-1* promoter because similar levels of specific transcripts were detected with a wt(−252) construct containing a transcription termination cassette (34) inserted upstream of the *mb-1* promoter (data not shown). An S1 nuclease assay which detects transcripts from the H4 reference gene ($H4^{ref}$) was used to control for transfection efficiency (lower panel) in this and subsequent transfection experiments. (C) Schematic diagram of the DNA probe used to detect the 5' ends of transcripts from the transfected *mb-1* promoter/β-globin gene constructs. nt, nucleotides.

ments into a blunt-ended *Sal*I β vector from which the 3.5-kb *Bam*HI fragment (containing the polyomavirus sequences) was removed. The −800βPy and −2800βPy constructs were generated by inserting a blunt-ended 840-bp *Hin*dIII fragment of the M9 bacteriophage (−770 to +70) and a 2.9-kb *Bam*HI-*Hin*dIII partial digestion fragment of the M9 bacteriophage (−2800 to +70), respectively, into a blunt-ended *Sal*I βPy vector.

Linker scanning mutants of the *mb-1* upstream factor (MUF) binding site region (Fig. 4A), the initiation site region (Fig. 8A), and the TATA box-containing mutant (Fig. 8B) were generated by site-directed mutagenesis of single-stranded DNA prepared from plasmid pBluMβ. pBluMβ was generated by subcloning a 450-bp *Bam*HI fragment of pwt(−252)βPy into Bluescript M13+ cut with *Bam*HI. Mutagenesis was performed as described previously (30), using the following oligonucleotide primers: 5'-CTCACTTCCTGT TCATCTGCAGCGGAGATGGGCTCC for LS(−66/−60), 5'-CAAACTCCGCCTCACTGCTGCAGCAGCCGTGCCGG AG for LS(−57/−51), 5'-GATTCCACTCCAAACTCACTG CAGCTTCCTGTTCAGCC for LS(−48/−42), 5'-GGTACC AGATTCCACGCTGCAGTCCGCCTCACTTCC for LS(−39/ −33), 5'-GTAGCCTTGAGGTACACTGCAGCACTCCAA ACTCCG for LS(−29/−23), 5'-GACCCCCTGGCATCCTC GAGCAGTGAGTCGGTTAG for LS(+20/+25), 5'-GACCC CCTGGCATCCTCGAGCAGTACTTCGGTTAGTTTGGG for LS(+13/+25), 5'-GGCATCGTCTCCCAGTACTTCGGT TAGTTTGGG for LS(+13/+15), and 5'-GCCTTGAGGTAC CAGTTTATACTCCAAACTCCGCC for TATA(−30). The deletion mutants Δ(−57/−42) and Δ(−57/−33) were generated by triple ligations of the 200-bp *Bam*HI-*Pst*I fragment of LS(−57/−51) with a *Bam*HI-*Hin*dIII Bluescript vector and either a 120-bp *Pst*I-*Hin*dIII fragment from LS(−48/−42) or a 110-bp *Pst*I-*Hin*dIII fragment from LS(−39/−33). The 12-bp insertion mutants LI(−18) and LI(+25) were generated by inserting an 8-bp *Sal*I linker into the filled-in *Asp* 718 site of LS(−66/−60) or the filled-in *Xho*I site of LS(+20/+25) in pBluMβ. The deletion mutant Δ(−18/+20) was generated by triple ligation of a Bluescript M13+ *Eco*RI-*Bam*HI vector with a 240-bp *Eco*RI-*Asp* 718 fragment (with the *Asp* 718 site blunt ended) and a 170-bp *Xho*I-*Bam*HI fragment (with the *Xho*I site blunt ended), both isolated from the LS(+20/+25) mutant in pBluMβ. All 14 mutant promoters were isolated from the Bluescript subclones as approximately 300-bp *Bam*HI-*Hin*dIII fragments, then blunt-ended with Klenow, and ligated into the blunt-ended *Sal*I βPy vector to produce the final constructs for transfection.

Plasmids MUFwt, MUFmut1, MUFmut2, and MUFmut3 were derivatives of the *fos*-chloramphenicol acetyltransferase (CAT) reporter plasmid p301-56 (17). Three copies of double-stranded oligonucleotides containing wild-type (wt) or mutant (mut) *mb-1* promoter sequences (−59 to −38) (Fig. 7A) were introduced in the same orientation into the *Sal*I site at −56 of the *fos*-CAT plasmid.

Cell culture and DNA transfection. NIH 3T3 and L cells were grown in Dulbecco modified Eagle medium containing 10% fetal bovine serum (Hyclone). Lymphoid cell lines were grown in RPMI containing 10% fetal bovine serum and 50 μM 2-mercaptoethanol.

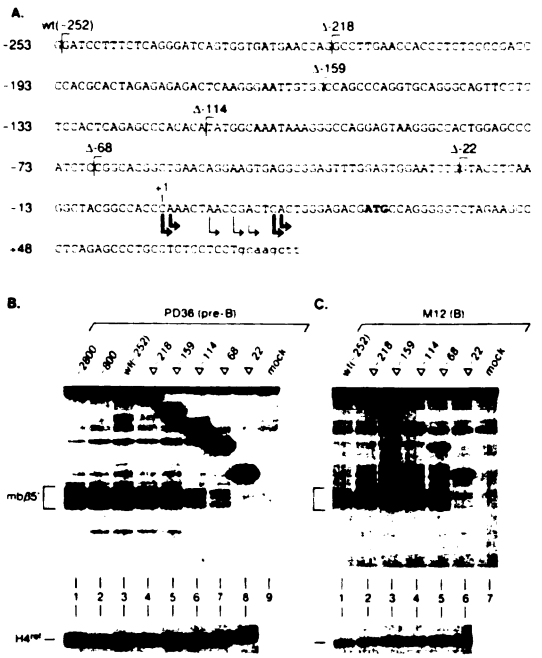


FIG. 3. Serial 5' end deletions define two functional regions of the *mb-1* promoter. (A) Nucleotide sequence of the coding strand of the *mb-1* gene promoter from −253 to +70 bp. The strong and weak transcription initiation sites are designated with thick and thin arrows, respectively. The +1 position is arbitrarily assigned to the 5'-most start site. Positions of the 5' ends of the various deletion mutants are indicated. (B) Determination of the relative number of specific transcripts derived from *mb-1* promoter 5'-deletion mutants in pre-B cells by S1 nuclease protection analysis. PD36 cells were transfected with gene constructs containing *mb-1* promoter fragments from the variable 5' boundary indicated (−2800, −800, −252, −218, −159, −114, −68, or −22 bp) to +70 bp. An S1 nuclease assay with RNA from untransfected PD36 cells was performed to control for background bands (lane 9). Specific transcripts from the *mb-1* promoter are designated mbβ5'. The strong bands of lower mobility represent vector read-through transcripts. Below, specific transcripts from the H4 reference gene ($H4^{ref}$) are shown. (C) Determination of the relative number of specific transcripts derived from *mb-1* promoter 5'-deletion mutants in B cells by S1 nuclease protection analysis. Because of their higher transfection efficiency, M12 B cells were transfected with gene constructs analogous to those in panel B but lacking the early region of polyomavirus (see Fig. 2A for structures of plasmids).

Transfections were done by the DEAE-dextran method (19), with 2 μg of DNA per ml. For cotransfections, this consisted of 1.5 μg of CAT reporter per ml plus 0.5 μg of pRSV-luciferase (15) per ml. Cells were harvested after 48 to 50 h, and total RNA (10) or extracts for luciferase and CAT assays (18) were prepared as described previously.

DNA probes. The template for generating the probe for endogenous *mb-1* transcripts (pMb1gc) was prepared by introducing a 255-bp *Stu*I-*Xba*I *mb-1* promoter fragment (from −217 to +38) into an *Eco*RV-*Xba*I-cut Bluescript subclone containing *mb-1* cDNA sequences from +38 to the poly(A) tract. To make the radiolabelled probe, a 393-bp *Nde*I-*Pvu*II fragment (Fig. 1C) was isolated from pMb1gc and 5' end labelled with ^{32}P, and the correct strand was isolated from a strand-separating gel as described previously (44). Single-stranded DNA probes for S1 nuclease mapping of the 5' ends of transcripts from the *mb-1* promoter constructs (Fig. 2C) were generated by extending a ^{32}P-5'-end-labelled primer comprising the mouse β-globin noncoding strand, second exon sequence beginning at the *Bam*HI site (27), 5'-GATCCACATGCAGCTTGTCACAG, using a single-stranded DNA template generated from the *mb-1* promoter/β-globin gene fusion subclone pBluMβ. *Eco*RI digestion of the extension product at a polylinker site was used to generate the 3' end of the probe. For the *mb-1* initiation site mutants, DNA probes were similarly prepared by using the same ^{32}P-labelled primer and templates containing the various mutations (Fig. 8A and B) in pBluMβ. The DNA probe used to detect endogenous H4 [H4(5')] and reference H4 transcripts was prepared as described previously (19).

Primer extension, S1 nuclease protection, luciferase, and CAT assays. Primer extension of endogenous mRNA was performed by using 1 μg of poly(A)$^+$ RNA, a ^{32}P-end-labelled complementary oligonucleotide (+63 to +43 of the *mb-1* sequence; see above), and murine leukemia virus reverse transcriptase (Bethesda Research Laboratories), using previously described methods (35).

The S1 nuclease protection assay was performed by using 20 μg of total RNA as described previously (19). Hybridization was at 48°C for 12 to 15 h. The hybrids were digested with 20 U of S1 nuclease (Pharmacia) at 37°C for 1 h. The protected DNA fragments were visualized by using 6% polyacrylamide-urea gels and autoradiography.

Luciferase and CAT assays were performed and quantitated as described previously (53) using a PhosphorImager (Molecular Dynamics).

Nuclear extracts and analysis of protein-DNA complexes. PD36 nuclear extract was prepared as described previously (16). Electrophoretic mobility shift and methylation interference assays were performed as described previously (22) except that $MgCl_2$ was not present in the binding reaction. The synthetic oligonucleotide probes and competitors used are described in the figure legends. Mobility supershift assays were performed by incubating PD36 nuclear extract with 1 μl of either normal rabbit serum or anti-Sp1 serum (a gift from Steve Jackson and Robert Tjian) for 30 min at room temperature before adding either the SV40(87/111) or the mb(−66/−23) oligonucleotide probe.

RESULTS

Early B-cell-specific transcription of the *mb-1* gene is initiated at multiple sites. The 5' ends of endogenous *mb-1* transcripts were mapped by using primer extension (Fig. 1A) and S1 nuclease protection (Fig. 1B) assays. Both methods detected multiple 5' ends of *mb-1* transcripts in poly(A)$^+$ RNA from the pre-B-cell lines 40E1, 38B9, and PD36 and the B-cell lines WEHI231 and M12 (Fig. 1A and B). Specific transcripts were not detected in NIH 3T3 fibroblast and J558L plasmacytoma cell lines, representing nonlymphoid and late-stage B cells, respectively. These data confirmed the early B-cell-specific expression pattern of the *mb-1* gene (43) and further indicated that transcription initiation is heterogeneous. The majority of *mb-1* transcripts initiated within a 16-bp region. The precise positions of the 5' ends were determined by electrophoresis of the primer extension products next to a sequencing ladder (Fig. 1A). For reference, we arbitrarily assigned +1 to the position of the 5'-most major initiation site.

S1 nuclease mapping of the *mb-1* RNA 5′ ends with a probe diagrammed in Fig. 1C revealed a similar set of major initiation sites and, in addition, minor initiation sites within a 32-bp upstream region (Fig. 1B, lanes 1 to 4). Together, these data indicate that transcription initiation from the *mb-1* promoter is heterogeneous but restricted primarily to a 16-bp region.

Lineage and differentiation stage specificity of the *mb-1* promoter. We examined the early B-cell specificity of *mb-1* gene transcription by transfecting various mouse cell lines with gene constructs containing *mb-1* promoter sequences linked to a β-globin reporter gene. The *mb-1* promoter fragments included sequences from either −2800 to +70 (−2800βPy) or −800 to −70 (−800βPy) (Fig. 2A). These gene constructs also contained polyomavirus sequences for replication in mammalian cells and an H4 reference gene to estimate transfection efficiency (19). Following transfection, total RNA was isolated and analyzed for the presence of specific *mb-1* transcripts by using an S1 nuclease protection assay (Fig. 2B). The DNA probe contains β-globin sequences and detects specific transcripts from the transfected *mb-1* promoter constructs but not from the endogenous *mb-1* gene (Fig. 2C). Accurately initiated transcripts from both the −2800βPy and the −800βPy *mb-1* promoter constructs were detected in RNA from transfected PD36 pre-B cells and M12 B cells (Fig. 2B, lanes 3 to 6). By contrast, specific transcripts were detected at a much lower abundance in RNA from the late-stage B cells J558L and SP2 (lanes 7 to 10), L fibroblasts (lanes 1 and 2), NIH 3T3 fibroblasts, and T-cell lines (data not shown). Transcription from the reference H4 transfection control gene was detected at a similar level in all RNA samples with the exception of PD36 cells, which were transfected less efficiently. These data indicate that *mb-1* promoter sequences from −800 to +70 contain regulatory information sufficient for pre-B and B-lymphocyte-specific transcription. The activity of the promoter, however, is approximately 10-fold higher in pre-B cells (PD36) than in B cells (M12).

Mapping of *mb-1* promoter elements. To delineate the sequences involved in *mb-1* promoter function, progressive 5′ deletions (Fig. 3A) were tested for their effects on promoter activity in PD36 pre-B cells (Fig. 3B). An *mb-1* promoter fragment comprising sequences from −252 to +70 showed the same activity as the larger promoter fragments (Fig. 3B, lanes 1 to 3) and was arbitrarily designated as wild type [wt(−252)]. Deletion of sequences 5′ to −159 had no significant effect on the generation of specific transcripts (lanes 3 to 5). Truncation of the *mb-1* promoter to −114 decreased its activity threefold. No further decrease was observed upon removal of additional sequences to −68 (lanes 5 to 7). However, truncation of the promoter to −22 decreased the level of expression 20-fold relative to that of the wt(−252) promoter (compare lanes 3 and 8). The strong bands of lower mobility represent vector read-through transcripts, because their migration is determined by the site of the 5′-deletion breakpoint. This 5′-deletion series localized functionally important sequences to the regions between −159 and −115 and between −68 and −23. Transfection of the 5′-deletion series into M12 B cells indicated that sequences upstream of −68 did not contribute to promoter function (Fig. 3C, lanes 1 to 6), suggesting a pre-B-cell-specific function of sequences between −159 and −115. When these truncated *mb-1* promoter/β-globin gene constructs were transfected into NIH 3T3 fibroblasts or J558L plasmacytoma cells, no activity could be detected with any of the promoter fragments (data not shown). Together, these

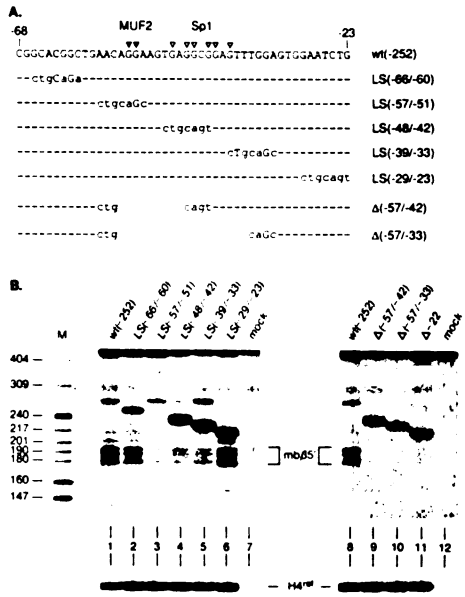


FIG. 4. Linker scanning mutagenesis of the *mb-1* proximal promoter region. (A) Nucleotide sequence of the coding strand of the *mb-1* promoter from −68 to −22 and of the linker scanning and deletion mutants. Dashes represent bases identical to the wild-type sequence. Mutant nucleotides, shown as lowercase letters and deletions, shown as gaps, were introduced into the wt(−252) *mb-1* promoter (containing sequences from −252 to +70). For the linker scanning mutants, coordinates in parentheses designate the region of nucleotide changes. For the internal deletion mutants, the coordinates designate the boundaries of the region of nucleotide changes and deletion. (B) Determination of the relative number of transcripts derived from the mutant *mb-1* promoters (shown in panel A) by S1 nuclease protection assay. The gene constructs were transfected into PD36 pre-B cells, and the transcripts were analyzed as described for Fig. 3. The strong bands of lower mobility represent cleavage by S1 nuclease at mismatches between vector read-through transcripts and the wild-type DNA probe. Below, the specific H4 reference ($H4^{ref}$) transcripts are shown.

data suggest that the *mb-1* promoter region residing between −68 and +70 is sufficient for early B-cell-specific activity.

To analyze the proximal *mb-1* promoter region (−66 to −23) in more detail, clustered point mutations (linker scanning mutations) were introduced in the wt(−252) promoter (Fig. 4A). The constructs were transfected into PD36 pre-B cells, and specific *mb-1* transcripts were detected by S1 nuclease protection analysis (Fig. 4B). The LS(−57/−51), LS(−48/−42), and LS(−39/−33) mutations each decreased the level of transcription fourfold relative to the wild-type promoter (Fig. 4B; compare lanes 3 to 5 with lane 1). Because these clustered mutations span a 25-bp region, the question arose as to whether they affect one or multiple factor binding sites. Therefore, two small internal deletions, Δ(−57/−42) and Δ(−57/−33), were introduced into the wt(−252) promoter (Fig. 4A). Both deletions decreased *mb-1* promoter activity 10- to 20-fold to a level comparable to that of the Δ-22 deletion (Fig. 4B, lanes 8 to 11). Together, these data suggest the presence of multiple factor binding sites in the *mb-1* proximal promoter region.

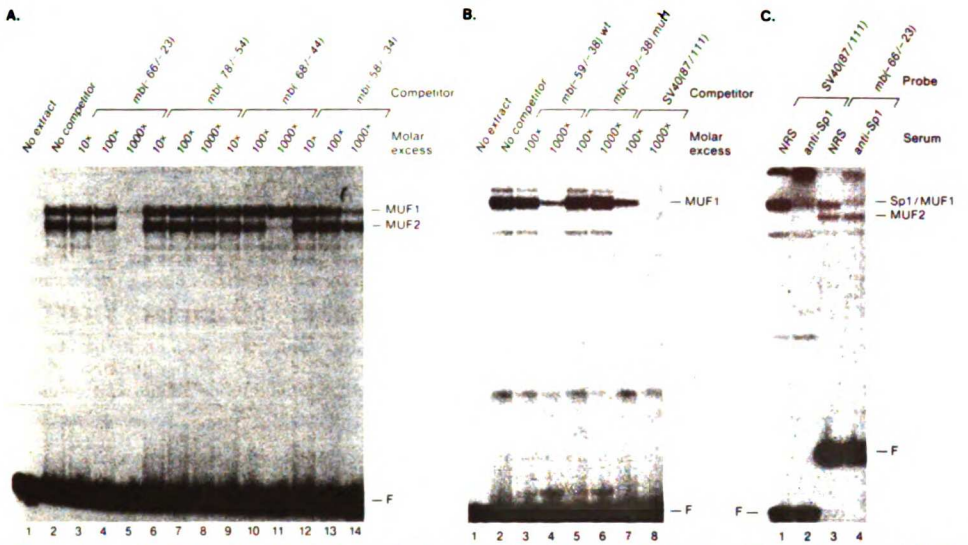


FIG. 5. Sequence-specific binding of nuclear factors to the functionally important proximal region of the *mb-1* promoter. (A) Detection of MUF1 and MUF2 by electrophoretic mobility shift assays. A double-stranded oligonucleotide probe containing *mb-1* promoter sequences from −66 to −23 was used in binding reactions with nuclear extract from PD36 pre-B cells. Two major complexes are observed with the addition of extract (lane 2). Both complexes are competed for with excess unlabelled probe oligonucleotide (lanes 3 to 5). To further localize sequences recognized by MUF1 and MUF2, binding competition was performed with three staggered 25-bp oligonucleotides spanning the proximal *mb-1* promoter from −78 to −54 (lanes 6 to 8), −68 to −44 (lanes 9 to 11), and −58 to −34 (lanes 12 to 14). (B) Specificity of MUF1 binding. A double-stranded oligonucleotide probe containing *mb-1* promoter sequences from −58 to −34 was incubated with PD36 pre-B-cell nuclear extract. The sequence specificity of MUF1 binding (lane 2) was determined by including excess wild-type (lanes 3 and 4) or mutant (lanes 5 and 6) oligonucleotide competitors. Both competitors contained *mb-1* promoter sequences from −59 to −38. The nucleotide changes in the mutant oligonucleotide are shown in Fig. 7A. The SV40(87/111) competitor contained a single Sp1 binding site from the SV40 early region (nucleotide positions 87 to 111) (4). (C) Evidence that MUF1 is identical to the transcription factor Sp1. Oligonucleotide probe containing a single Sp1 binding site from SV40 (lanes 1 and 2) or the MUF1 and MUF2 binding sites from the *mb-1* promoter (lanes 3 and 4) were added to PD36 pre-B-cell nuclear extract that had been preincubated with either normal rabbit serum (NRS) or rabbit anti-Sp1 serum (anti-Sp1). Protein-DNA complexes were separated in an electrophoretic mobility shift assay. F, free probe.

Factor binding to the proximal *mb-1* promoter region. We examined the binding of nuclear factors to the functionally important proximal *mb-1* promoter sequences by electrophoretic mobility shift assays with an end-labelled oligonucleotide probe (from −66 to −23) and a pre-B-cell nuclear extract. Two major protein-DNA complexes, termed mb-1 upstream factors 1 and 2 (MUF1 and MUF2), were detected (Fig. 5A, lane 2). The binding of both nuclear factors could be specifically competed for with excess unlabelled probe oligonucleotide (lanes 3 to 5) but not with an oligonucleotide comprising *mb-1* promoter sequences between −78 and −54 (lanes 6 to 8). The binding sites of both factors were further localized by competition with two overlapping 25-bp oligonucleotides spanning the proximal promoter region (lanes 9 to 14). These data indicate that MUF1 interacts with the −58 to −34 region and that MUF2 interacts with the −68 to −44 region.

To further characterize the MUF1-DNA complex, we used the shorter oligonucleotide probe (−58 to −34), which binds MUF1 only. The specificity of MUF1 binding to DNA was confirmed by competition with excess unlabelled wild-type or mutant *mb-1* oligonucleotide (Fig. 5B, lanes 2 to 6). The mutant oligonucleotide has the wild-type sequence GAGGCGGAG changed to GAGTATGAG. The wild-type sequence conforms to the reported consensus binding site for the transcription factor Sp1 (3). Therefore, we examined the relationship between MUF1 and Sp1 by using a known Sp1 binding site from the simian virus 40 (SV40) early region as a competitor for MUF1 binding (lanes 7 and 8). The Sp1 binding site from SV40 was a more efficient competitor for MUF1 binding than was the *mb-1* promoter sequence. To confirm that the MUF1 complex contains Sp1, nuclear extract was incubated with either normal rabbit serum or a polyclonal anti-Sp1 serum prior to the addition of the mb(−66/−23) oligonucleotide probe (Fig. 5C, lanes 3 and 4). The anti-Sp1 serum retarded the mobility of the MUF1-DNA complex but had no effect on the MUF2-DNA complex (lanes 3 and 4). Moreover, an Sp1-DNA complex formed on the SV40 oligonucleotide comigrated with the MUF1-DNA complex (lanes 1 and 3). As expected, anti-Sp1 serum retarded the mobility of this complex as well (lane 2). Together, these data strongly suggest that MUF1 is identical to Sp1.

Nucleotides in the proximal *mb-1* promoter which are contacted by MUF2 and Sp1 were determined by methylation interference analysis (Fig. 6A). Protein-DNA complexes were formed on partially methylated oligonucleotide probes comprising the binding site of either MUF2 or Sp1. On the coding strand, two adjacent G residues at −52 and −53 interfere with MUF2 binding when methylated, and all G residues within a region between −39 and −47 interfere with Sp1 binding when methylated (Fig. 6). On the noncoding

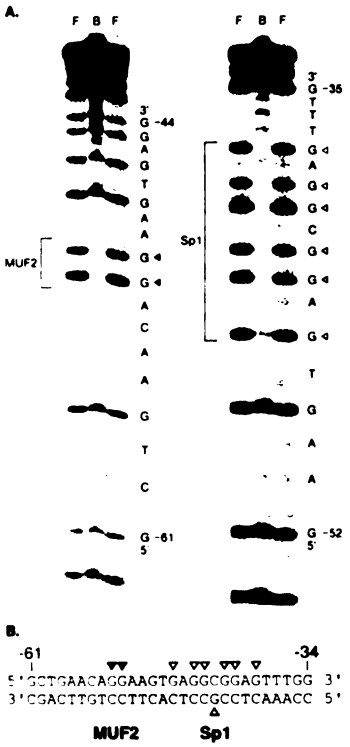


FIG. 6. Methylation interference analysis of the *mb-1* proximal promoter region binding MUF2 and Sp1. (A) Methylation interference analysis of MUF2 and Sp1 complexes to detect specific nucleotide contacts in the *mb-1* promoter. Partially methylated and end-labelled *mb-1* DNA probes from −68 to −39 and from −58 to −34 were incubated with PD36 nuclear extract to generate MUF2 complexes (left) or Sp1 complexes (right). Bound (B) and free (F) probe were separated in an electrophoretic mobility shift assay and processed as described previously (22). Nucleotide sequences of the upper (labelled) strands are shown for reference. The regions in which methylated G residues interfered with binding are indicated with brackets. (B) Summary of methylation interference data for both the upper (from panel A) and lower (data not shown) strands of the MUF2 and Sp1 binding sites. Arrowheads indicate nucleotides which interfere with DNA binding when methylated.

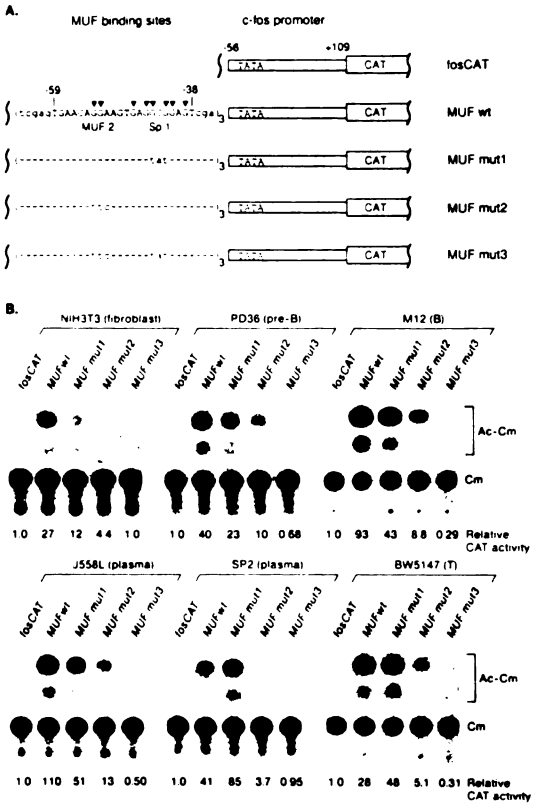


FIG. 7. Evidence that proximal promoter region factor binding sites augment expression from a minimal c-*fos* promoter. (A) Structure of the *mb-1*/c-*fos* promoter–CAT constructs. Three tandem copies of oligonucleotides containing wild-type or mutant *mb-1* promoter sequences from −59 to −38 were inserted into the *fos*-CAT reporter plasmid upstream of c-*fos* promoter sequences from −56 to +109. Dashes in the *mb-1* promoter sequences of MUFmut1, -2, and -3 represent bases identical to the MUFwt sequence, and lowercase letters indicate mutations. (B) Analysis of *mb-1*/c-*fos* promoter activity by CAT assays. The five constructs diagrammed in panel A were transfected into various cell lines along with the pRSV-luciferase construct (15) as an internal control. The amount of cell extract used in CAT assays was normalized to luciferase activity. Representative CAT assays, separating the acetylchloramphenicol (Ac-Cm) reaction products and chloramphenicol (Cm) by thin-layer chromatography, are shown. The average relative CAT activities were determined by multiple transfections and normalized to the CAT levels obtained with the *fos*-CAT construct.

strand, only a single G residue in the Sp1 site shows methylation interference, whereas no methylation interference could be detected in the MUF2 site (data not shown). These results indicate that MUF2 and Sp1 recognize adjacent and possibly nonoverlapping binding sites. Because no abundant protein-DNA complexes of lower mobility than MUF2 and Sp1 were observed in the gel mobility shift assay (Fig. 5A), their binding may be mutually exclusive.

Both MUF2 and Sp1 were detected by gel mobility shift assays in nuclear extracts from lymphoid and nonlymphoid cell lines, indicating a ubiquitous distribution (data not shown). Likewise, DNase I protection analysis of the proximal *mb-1* promoter identified a protected region from −59 to −38 with both lymphoid and nonlymphoid extracts (data not shown).

Binding sites for MUF2 and Sp1 can augment expression from a heterologous promoter. To examine whether the proximal *mb-1* promoter region between −59 and −38 can increase the activity of a heterologous promoter in a cell type-specific manner, three copies of oligonucleotides comprising either wild-type or mutant MUF2 and Sp1 sites were linked to the minimal c-*fos* promoter of a CAT reporter gene (Fig. 7A, fosCAT) (17). The chimeric promoter constructs were transfected into various B- and non-B-cell lines, together with the control plasmid pRSV-luciferase (15), and normalized CAT activity was determined after 48 h. The wild-type *mb-1* oligonucleotide containing both MUF2 and

Sp1 binding sites (MUFwt) increased the expression of the CAT gene in all transfected lymphoid and nonlymphoid cell lines 27- to 110-fold relative to that of the *fos*-CAT gene (Fig. 7B).

The individual contributions of the MUF2 and Sp1 sites to heterologous promoter activity were examined by introducing mutations in either or both factor binding sites (Fig. 7A). These mutations completely interfered with factor binding in vitro (Fig. 5B; data not shown for MUF2). Mutation of the Sp1 site (MUFmut1) decreased expression only twofold relative to the wild-type sequence in fibroblasts, pre-B, B, and J558L plasmacytoma cells; however, the same mutation increased expression twofold in SP2 plasmacytoma cells and in BW5147 T cells (Fig. 7B). By contrast, mutation of the MUF2 binding site (MUFmut2) consistently decreased CAT levels 4- to 10-fold in all cell types analyzed. The *mb-1* oligonucleotide containing mutations in both factor binding sites (MUFmut3) was unable to stimulate expression from the c-*fos* promoter.

A single copy of the various *mb-1* oligonucleotides had a similar effect on *fos* promoter activity, although the overall CAT levels were proportionally lower than those observed with three copies (data not shown). Taken together, the binding sites for MUF2 and/or Sp1 increased the expression from a heterologous promoter in both B- and non-B-cell lines. This ubiquitous activity of the proximal *mb-1* promoter region contrasts with its contribution to the early B-cell-specific function of the intact *mb-1* promoter.

The region of heterogeneous transcription initiation is determined by upstream sequences. In addition to its early B-cell-specific activity, the *mb-1* promoter differs from the Ig gene promoters in that it directs transcription initiation from multiple sites. In this respect, the *mb-1* promoter also differs from the TATA-less TdT promoter, which initiates transcription at a single nucleotide within the initiator element. The initiator was shown to be necessary and sufficient for specifying initiation from a single site (49, 50). Alignment of the 3'-most initiation site at +16 of the *mb-1* promoter with the initiator of the TdT promoter revealed a 5-of-7 match (CTCANTGG) and a 6-bp identity in the adjacent transcribed region (GGAGAC; Fig. 8A).

We examined the importance of these sequences for specifying transcription initiation by introducing clustered mutations in the wt(−252) *mb-1* promoter construct (Fig. 8A). For S1 nuclease analysis of RNA from transfected PD36 pre-B cells, each RNA sample was hybridized with a complementary probe containing the same base changes as the transfected gene construct (see Materials and Methods). The mutation LS(+13/+15) of the initiator homology and the LS(+13/+25) double mutation of both the initiator and transcribed-region homologies decreased the efficiency of utilization of the 3'-most pair of major initiation sites only by a factor of 2 and had no effect on the utilization of the initiation sites further upstream (Fig. 8A; Fig. 8C, lanes 3 and 4). The LS(+20/+25) mutation of the transcribed-region homology alone had no detectable effect on the utilization of multiple initiation sites (Fig. 8C, lane 2). Together, these data suggest that neither of these homologies in the *mb-1* promoter is required for the generation of the heterogeneous pattern of transcription initiation.

To examine the role of upstream or downstream regulatory sequences in the positioning of the multiple sites of transcription initiation, we inserted a 12-bp linker either at −18 or at +25 (Fig. 8B). The upstream insertion mutant LI(−18) utilized an additional set of initiation sites which extended 21 nucleotides 5' of +1. The utilization of the normal start sites (+1 to +16) was reduced threefold (Fig. 8C, lane 5). Thus, this mutation results in transcription initiation over a broader region compared with the wild type. Moreover, the center of the initiation site region is shifted 5' by approximately 12 nucleotides. In contrast, the downstream insertion mutant LI(+25) had a very small effect on the utilization of the RNA start sites, but, as expected, it extended all transcripts by 12 nucleotides (Fig. 8C, lane 6). Together, these results suggest that upstream sequences determine the region within which multiple sites can be used for transcription initiation.

This interpretation for the functional role of upstream sequences for the positioning of transcription start sites was supported by the analysis of another promoter deletion mutant, Δ(−18/+20), which lacks the region of normal transcription initiation. A novel set of initiation sites was utilized within a region downstream of the deletion breakpoint. The location of this novel initiation site region, relative to the upstream factor binding sites, corresponds to that of the wild-type promoter (Fig. 8B; Fig. 8C, lane 7). However, the pattern of multiple initiation sites of this mutant promoter is distinct from those of both the wild-type and LI(−18) mutant promoters, indicating that the nucleotide sequences within the region of transcription initiation determine the exact positions of the start sites.

To determine whether the heterogeneous initiation of transcripts could be attributed solely to the absence of a TATA element, we examined the effect of an insertion of the sequence TATAAA in the *mb-1* promoter between −31 and −26 [Fig. 8B, TATA(−30)]. The number of transcripts initiating from the predicted site downstream of the TATA element (+1), was increased threefold relative to that of the wild-type promoter. However, initiation from the other sites was still observed, although at a reduced level (Fig. 8C, lane 8). Thus, the multiple initiation sites in the *mb-1* promoter are utilized even in the presence of an introduced TATA element, suggesting an intrinsic TATA-independent mechanism of transcription initiation.

DISCUSSION

Role of ubiquitous factors in cell type-specific regulation. The function of the *mb-1* promoter in early B cells depends on sequences between −59 and −38, which bind the ubiquitous factors MUF2 and Sp1 in vitro. The proximal region of the *mb-1* promoter can also augment the activity of a heterologous minimal promoter. This increase in promoter activity is dependent upon both binding sites for MUF2 and Sp1 and can be detected in all analyzed lymphoid and nonlymphoid cell lines. The ubiquitous activity of these factor binding sites in a heterologous context contrasts with their functional role for the cell type-specific *mb-1* promoter. Although our current experimental data do not yet provide an insight into the mechanisms underlying the cell type-specific function of the *mb-1* promoter, several possible regulatory schemes can be considered.

First, negatively acting elements in the *mb-1* promoter may interfere with the function of the proximal promoter region in late-stage B cells and in non-B cells. Negative regulation has been implicated in the B-cell-specific function of the Ig heavy-chain enhancer (41). In this case, a putative negative regulator acting through the μE5 site can antagonize the positive function of both the ubiquitous factor ITF-1, which binds at the μE5 site, and the ubiquitous factor TFE3, which binds at the adjacent and nonoverlapping μE3 site (41). Because the minimal *mb-1* promoter fragment (−68

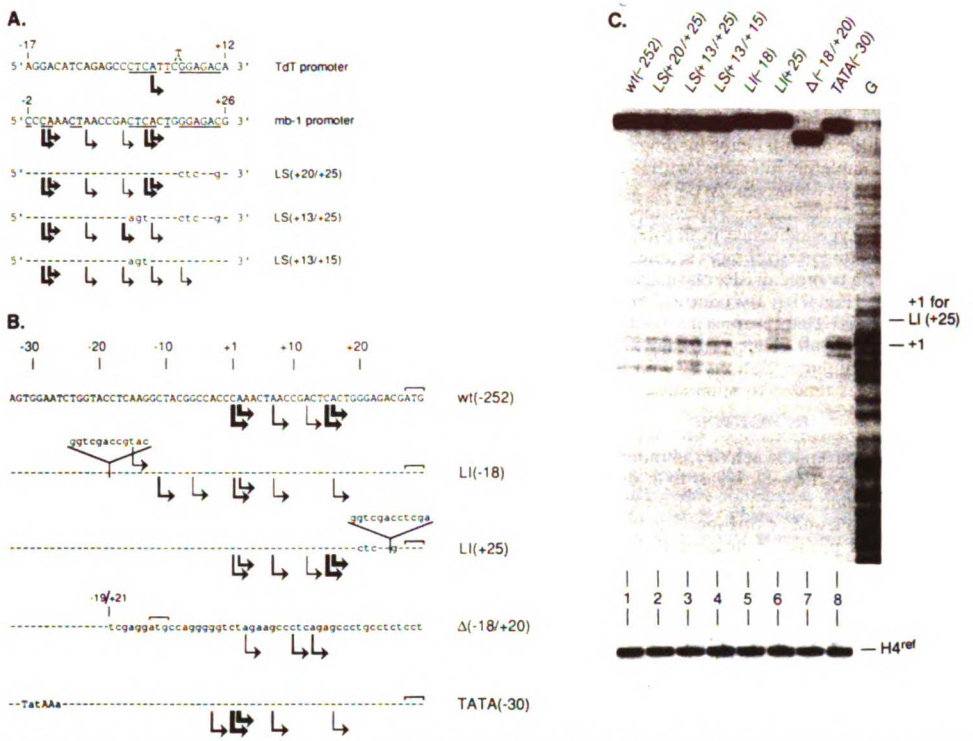


FIG. 8. Evidence that positioning of transcription initiation sites in the *mb-1* promoter is specified by upstream and initiation site sequences. (A) Aligned nucleotide sequences of the TdT and *mb-1* promoters and of three *mb-1* promoter mutants disrupting the homologies with the TdT promoter. Mutations were introduced into the wt(−252) construct containing *mb-1* promoter sequences from −252 to +70. The effect of each mutation (based on results shown in panel C, lanes 1 to 4) is summarized, with thick and thin arrows indicating the positions of major and minor initiation sites, respectively. (B) Nucleotide sequences of the initiation site regions of the wild-type and mutant *mb-1* promoters. LI(−18) and LI(+25) contain 12-bp linkers inserted at −18 and +25, respectively. Δ(−18/+20) has a deletion of sequences from −18 to +20. TATA(−30) has the sequence TATAAA substituted for sequences from −31 to −26. Thick and thin arrows indicate the positions of major and minor initiation sites, respectively (see panel C, lanes 5 to 8). (C) S1 nuclease protection analysis of RNA from PD36 pre-B cells transfected with constructs containing the various mutant *mb-1* promoters diagramed in panels A and B. Specific 5′ ends of *mb-1* transcripts were detected with single-stranded DNA probes similar to the one diagramed in Fig. 2C except that for each sample, a probe complementary to the specific mutant *mb-1* promoter sequence (shown in panel A or B) was used. The eight probes were generated by extending the same 5′-end-labelled oligonucleotide against the various mutant template DNAs (see Materials and Methods), ensuring they had identical specific radioactivity. +1 designates the predicted position of the 5′-most initiation site for all constructs except LI(+25) and Δ(−18/+20). +1 for LI(+25) designates the predicted position of the 5′-most initiation site for LI(+25).

to +70) displays the correct cell type-specific activity, a putative negatively acting element would most likely be present in the region downstream of −38. However, a preliminary analysis of deletion mutants spanning the region between −22 and +70 did not reveal any detectable promoter activity in non-B or late-stage B cells (data not shown).

A second possibility is that the ubiquitous activity of the proximal *mb-1* promoter region results from its artificial juxtaposition to the TATA-containing *fos* promoter. We consider it unlikely that the TATA element itself accounts for this loss of specificity because the insertion of a TATA element into the *mb-1* promoter does not abrogate its cell type specificity (data not shown). Nevertheless, juxtaposition of the proximal *mb-1* promoter to *fos* promoter sequences may generate a context in which either or both of the ubiquitous factors MUF2 and Sp1 can activate transcription in all cell types.

In a third regulatory scheme, MUF2 and/or Sp1 could participate in the cell type-specific *mb-1* promoter function by synergizing with an early B-cell-specific factor which binds to flanking promoter sequences. Consistent with such a model, we identified early B-cell factor (EBF), a nuclear factor which has the same expression pattern as the *mb-1* gene and binds to the distal *mb-1* promoter region between −181 and −160 (22). Because clustered point mutations within the EBF site decreased *mb-1* promoter activity in pre-B cells only threefold (22), EBF is unlikely to synergize with factors bound at the proximal promoter region.

Instead, we favor the view that the binding of MUF2 and Sp1 to *mb-1* promoter sequences in vitro and the ubiquitous activity of these factor binding sites in vivo do not necessarily reflect the participation of MUF2 and Sp1 in the regulation of the intact promoter. Possibly, the *mb-1* proximal promoter sequences are normally recognized by a less abundant cell type-specific factor, the binding of which is

favored in the context of the intact promoter by an interaction with another protein. The MUF2 binding site sequence shares seven contiguous base pairs (CAGGAAG) with the π box of the Ig heavy-chain enhancer, which binds a lymphoid-specific factor of unknown identity (32). Indeed, preliminary experiments indicate that the MUF2 binding site in the *mb-1* promoter can interact with the B-cell- and myeloid-specific factor PU.1 (26) present in crude nuclear extract (21a). This factor, together with EBF and possibly other cell type-specific factors, may participate in the early B-cell-specific regulation of the intact *mb-1* promoter.

Multiple initiation sites at a TATA-less promoter. The TATA-less *mb-1* promoter initiates transcription from multiple sites and thus differs from the TATA-less promoter of the TdT gene, which initiates transcription from a single site (49). Although two homologies (5 of 7 bp) with the TdT initiator are found in the *mb-1* promoter centered around the major start sites at +2 and +16, our data suggest that these initiation sites are functionally distinct from the TdT initiator. First, a mutation of the initiator homology around +16 of the *mb-1* promoter decreased expression from the 3'-most pair of initiation sites only by a factor of 2. Second, utilization of multiple initiation sites in the *mb-1* promoter is dependent on their location relative to upstream regulatory elements, whereas the TdT initiator can function autonomously (49, 50). The importance of upstream sequences in the approximate positioning of *mb-1* initiation sites is demonstrated by the analysis of two promoter mutations. An insertion of a 12-bp linker at −18 reduced initiation from the normal sites between +1 and +16 and activated additional initiation sites between −21 and +1. Moreover, an internal deletion removing the entire region of normal transcription initiation resulted in the use of a new set of initiation sites within a downstream region which was brought close to upstream promoter elements. The exact positioning of the multiple initiation sites in the *mb-1* promoter, however, is not determined by upstream sequences because the patterns of start sites differs among the wild-type and mutant promoters. Instead, the nucleotide sequences in the region of initiation appear to determine the exact transcription start sites. Taken together, these results suggest that sequences 5' of the initiation sites in the *mb-1* promoter specify a "window" in which multiple sites can be used for initiation. A similar model has been proposed for RNA polymerase II initiation in yeast cells (8).

Although the multiple initiation sites of the TATA-less *mb-1* promoter are reminiscent of the heterogeneous transcription observed with promoters that have a mutated TATA element (2, 20, 21), the molecular basis for the heterogeneity may be different. Mutations that generate a TATA element in the *mb-1* promoter at −30 favor initiation at a single site (+1); however, they do not eliminate initiation from downstream sites. Thus, the initiation of transcription at multiple sites appears to be an intrinsic property of the *mb-1* promoter and cannot be overridden simply by introducing a TATA element.

In the absence of a TATA element, what upstream element(s) and factor(s) determine the region of initiation in the *mb-1* promoter? Recent experiments using synthetic TATA-less promoters that contain an initiator and upstream Sp1 binding sites have shown that a TFIID fraction is required for transcriptional activation by Sp1 in vitro (40, 50). Experiments with highly fractionated TFIID suggested that an additional activity, termed a tethering factor, is required for activation of this TATA-less promoter. From these data, a model was proposed in which TFIID and the transcription complex are anchored through the tethering factor to Sp1 bound at the TATA-less promoter (40). The functionally important Sp1 site in the *mb-1* proximal promoter region could fulfill an analogous role and facilitate the formation of transcription complexes in the absence of a TATA element.

ACKNOWLEDGMENTS

We are grateful to Steve Jackson and Robert Tjian for providing the anti-Sp1 serum and to Michael Z. Gilman for the *fos*-CAT (p301-56) reporter plasmid. We thank Gail Mosely for assistance with manuscript preparation. We also thank Steve Hardy and Steve Yoshinaga for critical comments on the manuscript.

J.H. is a postdoctoral fellow of the Leukemia Society of America. This study was supported by funds from the Howard Hughes Medical Institute.

REFERENCES

1. **Blackwell, T. K., and F. W. Alt.** 1988. Immunoglobulin genes, p. 1–60. *In* B. D. Hames and D. M. Glover (ed.), Molecular immunology. IRL Press, Washington, D.C.
2. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
3. **Briggs, M. R., J. T. Kadonaga, S. P. Bell, and R. Tjian.** 1986. Purification and biochemical characterization of the promoter-specific transcription factor, Sp1. Science **234**:47–52.
4. **Buchman, A. R., L. Burnett, and P. Berg.** 1982. The SV40 nucleotide sequence, p. 799–841. In J. Tooze (ed.), DNA tumor viruses, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
5. **Cambier, J. C.** 1991. Capturing antigen receptor components. Curr. Biol. **1**:25–27.
6. **Campbell, K. S., and J. C. Cambier.** 1990. B lymphocyte antigen receptors (mIg) are non-covalently associated with a disulfide-linked, inducibly phosphorylated glycoprotein complex. EMBO J. **9**:441–448.
7. **Chao, M. V., P. Mellon, P. Charnay, T. Maniatis, and P. Axel.** 1983. The regulated expression of β-globin genes introduced into mouse erythroleukemia cells. Cell **32**:483–493.
8. **Chen, W., and K. Struhl.** 1985. Yeast mRNA initiation sites are determined primarily by specific sequences, not by the distance from the TATA element. EMBO J. **4**:3273–3280.
9. **Cherayil, B. J., and S. Pillai.** 1991. The ω/λ_4 surrogate immunoglobulin light chain is expressed on the surface of transitional B lymphocytes in murine bone marrow. J. Exp. Med. **173**:111–116.
10. **Chirgwin, T. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**:5294–5299.
11. **Clark, E. A., and P. J. L. Lane.** 1991. Regulation of human B-cell activation and adhesion. Annu. Rev. Immunol. **9**:97–127.
12. **Concino, M. F., R. F. Lee, J. P. Merryweather, and R. Weinmann.** 1984. The adenovirus major late promoter TATA box and initiation site are both necessary for transcription in vitro. Nucleic Acids Res. **12**:7423–7433.
13. **Cooper, M. D., and P. D. Burrows.** 1989. B-cell differentiation, p. 1–22. *In* T. Honjo, F. W. Alt, and T. H. Rabbits (ed.), Immunoglobulin genes. Academic Press, London.
14. **Davison, B. L., J.-M. Egly, E. R. Mulvihill, and P. Chambon.** 1983. Formation of stable preinitiation complexes between eukaryotic class B transcription factors and promoter sequences. Nature (London) **301**:680–686.
15. **DeWet, J. R., K. V. Wood, M. DeLuca, D. R. Helinski, and S. Subramani.** 1987. Firefly luciferase gene: structure and expression in mammalian cells. Mol. Cell. Biol. **7**:725–737.
16. **Dignam, J. D., R. M. Lebovitz, and R. G. Roeder.** 1983.

Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. **11**:1475–1489.
17. **Gilman, M. Z., R. N. Wilson, and R. A. Weinberg.** 1986. Multiple protein-binding sites in the 5'-flanking region regulate c-*fos* expression. Mol. Cell. Biol. **6**:4305–4316.
18. **Gorman, C. M., L. F. Moffat, and B. H. Howard.** 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. Mol. Cell. Biol. **2**:1044–1051.
19. **Grosschedl, R., and D. Baltimore.** 1985. Cell-type specificity of immunoglobulin gene expression is regulated by at least three DNA sequence elements. Cell **41**:885–897.
20. **Grosschedl, R., and M.L. Birnstiel.** 1980. Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants in vivo. Proc. Natl. Acad. Sci. USA **77**:1432–1436.
21. **Grosveld, G. C., C. K. Shewmaker, P. Jat, and R. A. Flavell.** 1981. Localization of DNA sequences necessary for transcription of the rabbit β-globin gene in vitro. Cell **25**:215–226.
21a. **Hagman, J.** Unpublished data.
22. **Hagman, J., A. Travis, and R. Grosschedl.** 1991. A novel lineage-specific nuclear factor regulates mb-1 gene transcription at the early stages of B cell differentiation. EMBO J., in press.
23. **Hermanson, G. G., M. Briskin, D. Sigman, and R. Wall.** 1989. Immunoglobulin enhancer and promoter motifs 5' of the B29 B-cell-specific gene. Proc. Natl. Acad. Sci. USA **86**:7341–7345.
24. **Hombach, J., T. Tsubata, L. Leclercq, H. Stappert, and M. Reth.** 1990. Molecular components of the B-cell antigen receptor complex of the IgM class. Nature (London) **343**:760–762.
25. **Kashiwamura, S.-I., T. Koyama, T. Matsuo, M. Steinmetz, M. Kimoto, and N. Sakaguchi.** 1990. Structure of the murine mb-1 gene encoding a putative sIgM-associated molecule. J. Immunol. **145**:337–343.
26. **Klemsz, M. J., S. R. McKercher, A. Celada, C. Van Beveren, and R. A. Maki.** 1990. The macrophage and B cell-specific transcription factor PU.1 is related to the *ets* oncogene. Cell **61**:113–124.
27. **Konkel, D. A., S. M. Tilghman, and P. Leder.** 1978. The sequence of the chromosomal mouse β-globin major gene: homologies in capping, splicing and poly A sites. Cell **15**:1125–1132.
28. **Kudo, A., and F. Melchers.** 1987. A second gene, V_{preB} in the λ5 locus of the mouse, which appears to be selectively expressed in pre-B lymphocytes. EMBO J. **6**:2267–2272.
29. **Kudo, A., N. Sakaguchi, and F. Melchers.** 1987. Organization of the murine Ig-related λ_5 gene transcribed selectively in pre-B lymphocytes. EMBO J. **6**:103–107.
30. **Kunkel, T. A.** 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc. Natl. Acad. Sci. USA **82**:488–492.
31. **Landau, N. R., T. P. St. John, I. L. Weissman, S. C. Wolf, A. E. Silverstone, and D. Baltimore.** 1984. Cloning of terminal transferase cDNA by antibody screening. Proc. Natl. Acad. Sci. USA **81**:5836–5840.
32. **Libermann, T. A., and D. Baltimore.** 1991. Transcriptional regulation of immunoglobulin gene expression. Mol. Aspects Cell. Regul. **6**:385–407.
33. **Mathis, D. J., and P. Chambon.** 1981. The SV40 early region TATA box is required for accurate in vitro initiation of transcription. Nature (London) **290**:310–315.
34. **Maxwell, I. H., G. S. Harrison, W. M. Wood, and F. Maxwell.** 1989. A DNA cassette containing a trimerized SV40 polyadenylation signal which efficiently blocks spurious plasmid-initiated transcription. BioTechniques **7**:276–280.
35. **McKnight, S. L., and R. Kingsbury.** 1982. Transcriptional control signals of a eukaryotic protein-coding gene. Science **217**:316–324.
36. **Muller, M. M., S. Ruppsert, W. Schaffner, and P. Matthias.** 1988. A cloned octamer transcription factor stimulates transcription from lymphoid-specific promoters in non-B cells. Nature (London) **336**:544–551.
37. **Myers, R. M., K. Tilly, and T. Maniatis.** 1986. Fine structure genetic analysis of a β-globin promoter. Science **232**:613–618.
38. **Nakajima, N., M. Horikoshi, and R. G. Roeder.** 1988. Factors involved in specific transcription by mammalian RNA polymerase II: purification, genetic specificity, and TATA box-promoter interactions of TFIID. Mol. Cell. Biol. **8**:4028–4040.
39. **Oettinger, M. A., D. G. Schatz, C. Gorka, and D. Baltimore.** 1990. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. Science **248**:1517–1523.
40. **Pugh, B. F., and R. Tjian.** 1990. Mechanism of transcriptional activation by Sp1: evidence for coactivators. Cell **61**:1187–1197.
41. **Ruezinsky, D., H. Beckmann, and T. Kadesch.** 1991. Modulation of the IgH enhancer's cell type specificity through a genetic switch. Genes Dev. **5**:29–37.
42. **Sakaguchi, N., C. N. Berger, and F. Melchers.** 1986. Isolation of a cDNA copy of an RNA species expressed in murine pre-B cells. EMBO J. **5**:2139–2147.
43. **Sakaguchi, N., S.-I. Kashiwamura, M. Kimoto, P. Thalmann, and F. Melchers.** 1988. B lymphocyte lineage-restricted expression of *mb-1*, a gene with CD3-like structural properties. EMBO J. **7**:3457–3464.
44. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
45. **Sawadogo, M., and R. G. Roeder.** 1985. Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. Cell **43**:165–175.
46. **Schatz, D. G., M. A. Oettinger, and D. Baltimore.** 1989. The V(D)J recombination activating gene, RAG-1. Cell **59**:1035–1048.
47. **Scheidereit, C., J. A. Cromlish, T. Gerster, K. Kawakami, C. Balmaceda, R. A. Currie, and R. G. Roeder.** 1988. A human lymphoid-specific transcription factor that activates immunoglobulin genes is a homeo-box protein. Nature (London) **336**:551–557.
48. **Singh, H., J. H. LeBowitz, A. S. Baldwin, and P. A. Sharp.** 1986. A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. Nature (London) **319**:154–158.
49. **Smale, S. T., and D. Baltimore.** 1989. The "initiator" as a transcription control element. Cell **57**:103–113.
50. **Smale, S. T., M. C. Schmidt, A. J. Berk, and D. Baltimore.** 1990. Transcriptional activation by Sp1 as directed through TATA or initiator: specific requirement for mammalian transcription factor IID. Proc. Natl. Acad. Sci. USA **87**:4509–4513.
51. **Staudt, L. M., R. G. Clerc, H. Singh, J. H. LeBowitz, P. A. Sharp, and D. Baltimore.** 1988. Cloning of a lymphoid-specific cDNA encoding a protein binding the regulatory octamer DNA motif. Science **241**:577–580.
52. **Tedder, T. F., and C. M. Isaacs.** 1989. Isolation of cDNAs encoding the CD19 antigen of human and mouse B lymphocytes. A new member of the immunoglobulin superfamily. J. Immunol. **143**:712–717.
52a. **Travis, A.** Unpublished data.
53. **Travis, A., A. Amsterdam, C. Belanger, and R. Grosschedl.** 1991. LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor α enhancer function. Genes Dev. **5**:880–894.
54. **Wasylyk, B., R. Derbyshire, A. Guy, D. Molko, A. Roget, R. Teoule, and P. Chambon.** 1980. Specific *in vitro* transcription of conalbumin gene is drastically decreased by single-point mutation in TATA box homology sequence. Proc. Natl. Acad. Sci. USA **77**:7024–7028.
55. **Wirth, T., L. Staudt, and D. Baltimore.** 1987. An octamer oligonucleotide upstream of a TATA motif is sufficient for lymphoid-specific promoter activity. Nature (London) **329**:174–178.

CHAPTER FOUR:

Purification of EBF, a novel early B cell-specific nuclear factor important to *mb-1* promoter function.

ABSTRACT

EBF is a nuclear factor that recognizes a functionally important sequence element in the promoter of the gene that encodes the immunoglobulin-associated protein Mb-1. Like the *mb-1* gene, EBF is expressed specifically in early B lymphocytes. We have purified EBF by sequence-specific DNA affinity chromatography. A 60 kD polypeptide species copurified with EBF DNA binding activity. Renaturation of this polypeptide after isolation from a sodium dodecyl sulfate polyacrylamide gel reconstituted EBF DNA binding activity. Affinity purified EBF protected the same region of the mb-1 promoter from DNase I digestion as EBF from crude nuclear extract. Photoaffinity cross-linking of purified EBF resulted in the formation of a single protein-DNA adduct of 80 kD, consistant with the formation of a binary complex between the 60 kD polypeptide and the 30 basepair oligonucleotide.

INTRODUCTION

Studies of the early B lymphocyte-specific *mb-1* promoter identified an early B cell-specific nuclear factor (EBF) which recognizes a sequence element in the distal promoter between positions -160 and -180 (Hagman et al., 1991). Mutation of the EBF binding site was found to have a three- to four-fold effect on mb-1 promoter function in transient transfection assays of pre-B cells. In the context of flanking sequences, the EBF binding site was able to enhance expression from a heterologous promoter in early B cells, but not in T cells or fibroblasts, which do not express EBF (Hagman et al., 1991).

EBF is distinct from other known nuclear factors that are preferentially expressed in B lymphocytes. These include Oct-2, which recognizes the OCTA site in the regulatory sequences of immunoglobulin genes and is expressed in B cells and some T cells (Muller et al., 1988; Scheidereit et al., 1988; Staudt et al., 1988); PU.1, which binds an element in the promoter of the major histocompatibility complex (MHC) class II gene I-Aβ and is expressed in B cells and macrophages (Klemsz et al., 1990); and BSAP, which is expressed in early B cells and was identified as binding to a sequence element in sea urchin late histone promoters (Barberis et al., 1990). Athough EBF and BSAP have identical expression patterns during B cell differentiation, they were shown to be distinct factors based on both binding site competition and photoaffinity cross-linking experiments (Hagman et al., 1991).

Here, we describe the purification of EBF from pre-B cell nuclear extract using sequence-specific DNA affinity chromatography (Kadonoga and Tjian, 1986). Purified EBF and EBF in crude nuclear extract generate protein-DNA complexes of the same size in electrophoretic mobility shift assays and protect the same region in DNase I footprint assays. A 60 kD polypeptide species was concluded to represent

EBF on the basis of renaturation of EBF activity after SDS gel electrophoresis. Photoaffinity cross-linking data supports this conclusion.

MATERIALS AND METHODS

Assays of protein-DNA complexes

Electrophoretic mobility shift and DNase I footprint assays were performed essentially as described (Hagman et al., 1991) with the following modifications. The DNA probe used in the mobility shift assays was a 121 bp restriction fragment containing mb-1 promoter sequences from -252 to -141 (coordinates as in Travis et al., 1991) which was dephosphorylated with CIP and ^{32}P-labelled using T4 polynucleotide kinase. For affinity purified EBF, 100 µg/ml of poly(dIdC)-poly(dIdC) was used in the mobility shift assays and 2 µg/ml was used in the footprint analysis. The wild type and mutant specific competitors are the wt and m4 oligonucleotides given in Hagman et al., 1991.

S-Sepharose and DNA affinity column chromatography

Nuclear extracts were prepared essentially as described (Dignam et al., 1983) from 38B9 pre-B cells grown in RPMI supplemented with 10% equine serum. The undialized nuclear extract was diluted to 0.1 M NaCl with HM buffer (20mM Hepes (pH 7.9), 20% glycerol, 2.5 mM $MgCl_2$, and 1mM DTT) and chromatographed on an S-Sepharose (fast flow) matrix in an SR10/50 column (Pharmacia). Bound protein was eluted with a linear gradient from 0.1 to 0.6 M KCl. Peak fractions of EBF activity were determined by mobility shift assays, pooled and dialyzed against 80 mM KCl in HM buffer.

To generate a DNA-affinity matrix (Kadonaga and Tjian, 1986) specific to EBF, the oligonucleotides 5'-GCCAGATATCCTCAAGGGAATTGTG and 5'-

TATCTGGCCACAATTCCCTTGAGGA, which constitute when hybridized the m1 mutant version of the EBF binding site from the mb-1 promoter (Hagman et al., 1991) with eight base over-hanging complementary ends, were self-ligated to generate concatamers (average length ~20 copies). They were then coupled to CNBr-activated Sepharose CL-2B (Pharmacia). The dialyzed, S-Sepharose-fractionated EBF was mixed with sheared poly(dIdC)-poly(dIdC) (average size ~500 bp) to 20 μg/ml and applied to two 1.0 ml DNA affinity columns. The columns were washed in three steps with HM buffer containing 0.30, 0.35, and 0.30 M KCl and EBF activity eluted with HM buffer containing 1 M KCl. The eluate was dialized to 80 mM KCl, mixed with poly(dIdC)-poly(dIdC) to 2 μg/ml and re-applied to another 1.0 ml DNA affinity column and washed and eluted as described above.

Renaturation of gel purified EBF

Samples of approximately 40 and 400 ng affinity purified EBF were applied to adjacent lanes of a 10% SDS-polyacrylamide gel. After electrophoresis, the gel was cut between the two lanes, notches were made for alignment of the two pieces, and the piece containing 40 ng was silver stained. Segments of the unstained gel lane corresponding to the 60 and 110 kilodalton bands were excised and processed as described (Hager and Burgess, 1980) with the following modifications. Protein was eluted from the crushed gel slice with 1 ml elution buffer (20mM Hepes 7.9, 0.1% SDS, 5mM DTT, 0.1 mg/ml BSA, and 200mM NaCl). The acetone precipitate was resuspended in 6 μl of 6 M guanidine HCl in dilution buffer (20mM Hepes 7.9, 20% glycerol, 2.5mM $MgCl_2$, 0.5μM $ZnCl_2$, 3mM DTT, 0.1 mg/ml BSA, and 150mM KCl). After fifty-fold dilution of the guanidine HCl, the samples were incubated for 2 h at RT and then assayed for EBF activity.

RESULTS

Purification of EBF

The purification of EBF is summarized in Table 1. EBF DNA binding activity was measured using an electrophoretic mobility shift assay with a labelled fragment of the mb-1 promoter containing the EBF binding site. Crude nuclear extract was prepared from the EBF-expressing pre-B cell line 38B9. For the first chromatographic step, the nuclear extract was diluted to 0.1 M NaCl and applied to a sulfomethyl (S) Sepharose cation exchange column. A linear gradient of KCl (0.1 to 0.6 M) was applied and EBF activity was found to elute in the 0.3 to 0.4 M fractions. The EBF-containing fractions were pooled, dialized against buffer containing 80 mM KCl and applied to a DNA affinity column containing concatamers of a high affinity EBF binding site. After washing the column, EBF activity was eluted with 1 M KCl and dialized to 80 mM KCl. This material was then chromatographed a second time on a DNA affinity column.

The DNA affinity column fractions were examined by SDS gel electrophoresis. Two major polypeptide species of 110 and 60 kD copurified with the EBF activity (Figure 1A). The 60 kD species consistantly ran as a diffuse band from 58 to 62 kD.

Renaturation of EBF

To determine if one or both of the polypeptide species purified using the DNA affinity column corresponded to EBF, the polypeptides were separated by SDS gel electrophoresis and eluted from the gel. The 115 and 60 kD species (p115 and p60) were precipitated with acetone to remove SDS and renatured through the use of 6 M guanidine hydrochloride (Hager and Burgess, 1980). The gel purified, renatured proteins were examined with an electrophoretic mobility shift assay

Table 1. Purification of EBF.

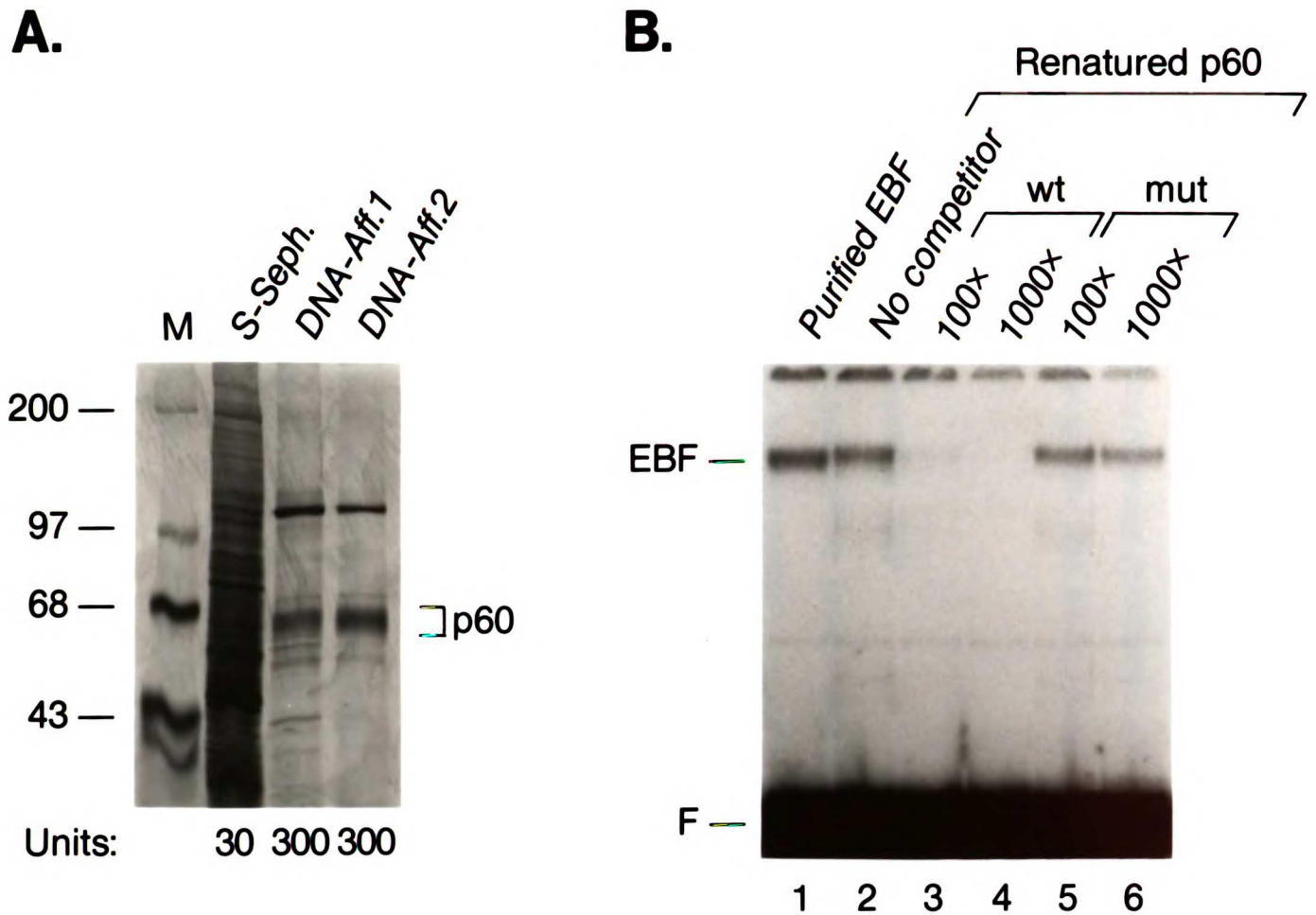
Procedure	Total protein (mg)	Activity total (units*)	Activity specific (units/mg)	Purification factor	Cummulative yield (%)
38B9 pre-B cells (~120g)	5900				
nuclear extract	550	700,000	1300		100
S Sepharose	50	270,000	5400	4x	39
EBF-site Sepharose	0.045	90,000	2,000,000	370x	13
EBF-site Sepharose	0.018	71,000	4,000,000	2x	10

* One unit of activity is defined as the amount of EBF required in a electrophoretic mobility shift assay to bind 0.1 out of 0.5 fmol of a 120-bp probe containing the EBF site from the mb-1 promoter.

Figure 1. Polypeptide composition of EBF.

(A) Silver staining of polypeptides present in EBF-containing fractions at different purification stages (S-Sepharose and DNA affinity columns 1 and 2). Numbers under the lanes represent the amount of EBF DNA binding activity (in units) applied to each lane. p60 represents the region of gel excised for eluting and renaturing.

(B) Renaturation of EBF DNA binding activity from gel purified p60. Samples of native EBF (lane 1) or gel purified and renatured p60 (lanes 2-6) were examined by an electrophoretic mobility shift assay. The bands representing the EBF-DNA complex (EBF) and free DNA (F) are labelled. The sequence specificity of DNA binding by renatured p60 (lane 2) was determined by including excess wild-type (lanes 3 and 4) or mutant (lanes 5 and 6) oligonucleotide competitors. Both competitiors contained mb-1 promoter sequences from -154 to -183, but the mutant competitor had nucleotide changes which abrogate EBF binding.


A.
M
S-Seph.
DNA-Aff.1
DNA-Aff.2
200
97
68
43
p60
Units:
30
300
300
B.
Renatured p60
Purified EBF
No competitor
wt
mut
100x
1000x
100x
1000x
EBF
F
1
2
3
4
5
6

using an end-labelled fragment from the mb-1 promoter containing the EBF binding site. Renatured p60 produced a protein-DNA complex which comigrated with the complex formed with native, purified EBF (Figure 1B, lanes 1 and 2). Moreover, the binding of renatured p60 could be specifically competed with a wildtype but not a mutant oligonucleotide competitor comprising the EBF binding site (lanes 3-6). No DNA binding activity was found for renatured p115 (data not shown). These data suggest that p115 represents a non-specific DNA binding protein contaminant and that p60 represents EBF.

Biochemical properties of purified EBF

A previous analysis of EBF in crude nuclear extract using photoaffinity cross-linking suggested that EBF contains two polypeptide components that contact DNA (Hagman et al., 1991). This experiment was repeated using affinity purified EBF for photoaffinity cross-linking in solution. An internally radiolabelled and 5-bromo-deoxyuridine substituted probe comprising the EBF binding site of the mb-1 promoter was incubated with affinity purified EBF and then irradiated with UV light (300nm) to covalently cross-link EBF to the probe DNA. The DNA-protein adducts were then resolved by SDS gel electrophoresis and visualized by autoradiography (Figure 2). A single major species of ~80 kilodaltons was detected, suggesting that only one EBF polypeptide crosslinks to the probe under these conditions. Formation of this DNA-protein adduct was specifically ingibited by a molar excess of wildtype but not mutant oligonucleotide competitor comprising the EBF binding site (data not shown). The larger size of the cross-linked polypeptide relative to p60 (Figure 1A) is consistant with the additional molecular weight of the cross-linked oligonucleotide.

To determine if affinity purified EBF binds and protects the same region of the mb-1 promoter as EBF present in crude nuclear extract, a DNase I protection

Figure 2. Photoaffinity cross-linking of EBF to DNA. An internally 32P-labelled and 5-bromo-deoxyuridine substituted probe comprising the ENF binding site was incubated with affinity purified EBF and then irradiated with UV light for 5 or 10 minutes as indicated to covalently cross-link EBF to the DNA probe. The protein-DNA adducts were ressolved by SDS gel electophoresis and visualized by autoradiography. The positions and sizes (in kilodaltons) of the molecular weight markers are indicated (M).

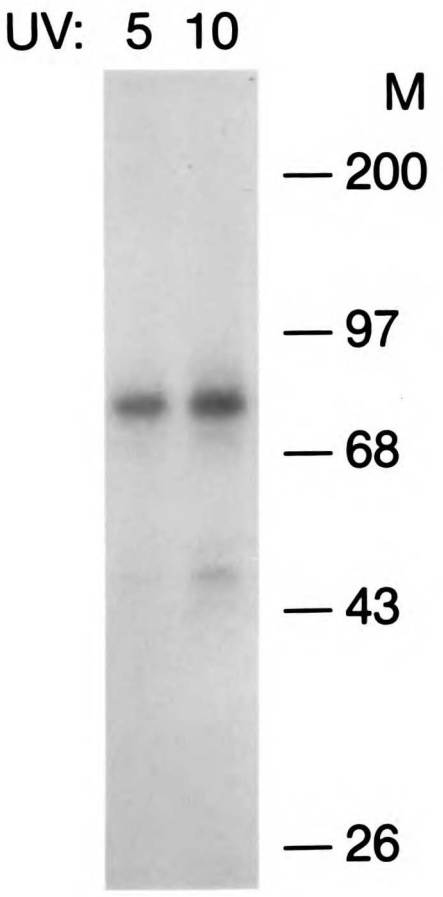
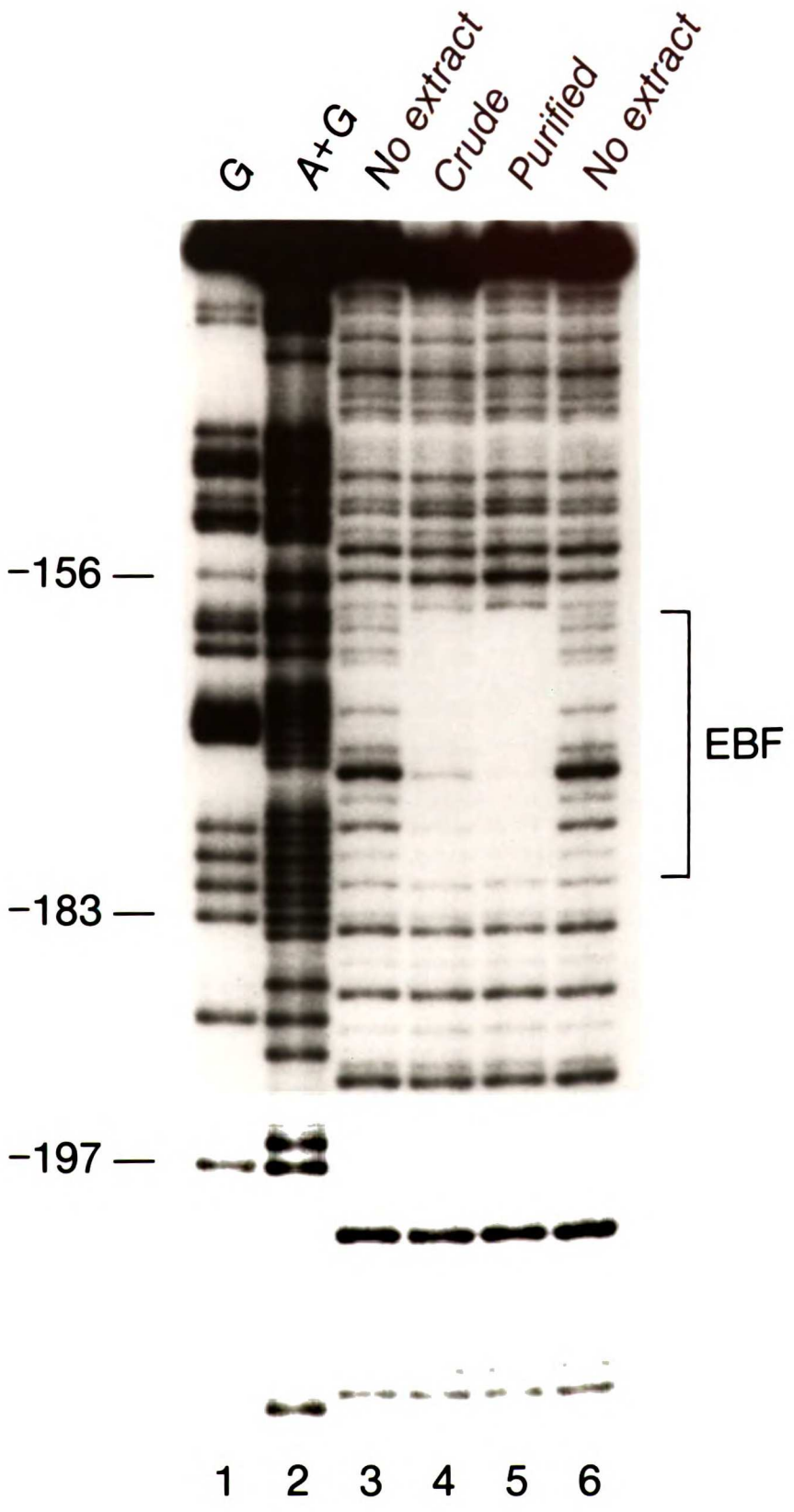

UV: 5 10
M
200
97
68
43
26

Figure 3. DNase I footprint protection of the mb-1 promoter with affinity purified EBF. The coding strand of a distal mb-1 promoter fragment (-252 to -113) was 5' end-labelled with 32P, incubated with either crude nuclear extract (lane 4) or affinity purified EBF (lane 5), digested with DNase I, and separated by denaturing gel electrophoresis. Lane 3 and 6 are negative controls with no EBF added. Lane 1 and 2 represent G and A+G Maxam-Gilbert sequencing ladders. Numbers on the left are mb-1 promoter coordinates (Travis et al., 1991b). The bracket indicates the extent of protection from DNase I cleavage.


G
A+G
No extract
Crude
Purified
No extract
−156
−183
−197
EBF
1
2
3
4
5
6

(footprinting) experiment was performed with a labelled DNA fragment comprising the mb-1 promoter from -252 to -113 (Figure 3). As previously shown (Hagman et al., 1991), EBF from crude nuclear extract protects a region of the coding strand of the mb-1 promoter from -160 to -179 (Figure 3, lane 4) from DNase I digestion. Affinity purified EBF protects the same region (lane 5), suggesting that this 20 bp footprint is not due to the cooperative binding of EBF with some other factor present in crude nuclear extract.

DISCUSSION

The early B lymphocyte-specific activity of the mb-1 promoter is dependent in part on the presence of a functional binding site for EBF, a novel early B cell-specific nuclear factor (Hagman et al., 1991). In order to further characterize EBF and its interaction with the mb-1 promoter, we have used sequence-specific DNA affinity chromatography to purify it from pre-B cell nuclei. We have identified a 60 kD polypeptide species which copurifies with EBF and has EBF DNA binding activity when renatured after extraction from an SDS gel slice.

Consistant with purified EBF being identical to EBF identified in crude nuclear extract, crude and purified EBF were found to generate identical DNase I footprints on the distal mb-1 promoter. However, in contrast to previous results suggesting that EBF is composed of two polypeptide chains (Hagman et al., 1991), we find that purified EBF generates a single protein-DNA adduct upon photoaffinity cross-linking. The basis of this discrepancy could be an artifact of the use of crude nuclear extract in previous cross-linking experiments.

Amino acid sequence data has recently been obtained using the purified 60 kD EBF polypeptide. Comparision of this sequence with sequences in the Protein

Identification Resource/NBRF has indicated that this protein is novel. The amino acid sequence was used to derive degenerate oligonucleotides which were used to generate an EBF cDNA fragment by polymerase chain reaction. RNA blot analysis using this fragment as the probe has confirmed that the gene encoding the 60 kD protein is expressed specifically in early B lymphocytes (J. Hagman, unpublished results).

SUMMARY AND PERSPECTIVES

This work describes the isolation of putative regulators of lymphoid-specific gene expression and differentiation by using two different approaches. One was an established approach for isolating mammalian regulators involving the purification of a cell type-specific nuclear factor that binds to a regulatory element in a cell type-specific promoter. The other approach was novel, involving the isolation of several lymphoid-specific cDNA clones, followed by the identification of those cDNA clones encoding polypeptides with DNA binding activity. The established approach has the advantage that a cell type-specific target gene and binding site are in hand prior to the isolation of the putative regulator and can be used for functional and biochemical studies. The novel approach has the advantage that it does not rely on DNA binding assays to identify a cell type-specific factor. This avoids the potential problem of a ubiquitous factor binding to the same site and obscuring the identification of a cell type-specific factor.

Both approaches are limited in scope to those regulators of lymphoid-specific gene expression which are themselves lymphoid-specific and which bind DNA. In support of the notion that this subset of regulators is important, several mammalian cell type-specific factors have been identified that bind to the regulatory sequences of cell type-specific structural genes and appear to activate their transcription (see Introduction, Table 1). Nevertheless, it is likely that the early-acting regulators of cell type specification are not exclusively cell type-specific DNA binding proteins. Although the early-acting regulators which specify the lineages of the definitive mammalian embryo are unknown, consideration of the early events of mammalian embryogenesis suggests certain likely features of these regulators.

The specification of most, if not all mammalian cell types appears to depend on the particular extracellular contexts of their lineage founder cells (Davidson, 1990). This suggests that ligand-receptor interactions at the founder cell surface initiate intracellular signals which activate the expression of particular cell type-specific regulators. A given lineage founder cell may thus have a range of possible fates defined by the range of potentially functional cell type-specific regulators. Activation of a particular regulator (or set of regulators) as a result of intercellular interactions starts the founder cell down a particular course of cell type-specific differentiation. In this model, the lineages chosen by various founder cells depend on the particular extracellular signals they receive. In addition to this "conditional specification" of cell lineages, the developmental programs of organisms from a large number of animal taxa depend on the "autonomous specification" of cell lineages. In this case, maternal factors sequestered in particular regions of the egg are inherited by particular founder cells and then specify particular lineages autonomously. Autonomous specification of cell lineages is particularly extensive in *Drosophila* embryogenesis, where a spatially organized molecular structure along both the anterior-posterior and dorsal-ventral axes of the egg is established during oogenesis. In mammalian embryogenesis however, no axial organization of the egg is known to be involved, and all lineages are believed to be specified conditionally (Davidson, 1990).

Conditional specification implies that a pluripotent founder cell must have a greater range of potentially functional cell type-specific regulators than will be used to specify a given fate. In addition, it must express the necessary diversity of ligand receptors and signal transduction machinery to mediate the response to several different potential extracellular signals. The ligands themselves would be more spatially confined to particular cellular positions in the developing embryo than would be the receptors (Davidson, 1990). Thus the fate of a given founder cell

would depend on the expression of particular ligands by adjacent cells in the embryo. The specification of a particular cell type might include, in addition to the expression of a battery of cell type-specific structural genes, the expression of a new set of spatially confined ligands (Davidson, 1990). In this way, as the cells of the embryo begin to diversify, they could establish new spatial patterns of ligand expression, which in turn could contribute to the further diversification of cell types.

From this perspective, important regulators of cell type-specific gene expression would include specific ligands, ligand receptors, and signal transduction machinery components, all acting upstream of the cell type-specific DNA binding proteins and the battery of cell type-specific structural genes that ultimately define a given cell type. For the hematopoietic lineages, many candidate ligands and ligand receptors involved in differentiation already exist among the large families of interleukins and interleukin receptors. By working backwards from the regulatory sequences of both cell type-specific structural and regulatory genes, it should be possible to determine which molecules actually participate in the early events that specify the hematopoietic lineages.

Although several lymphoid-specific DNA binding factors that recognize the regulatory sequences of lymphoid-specific structural genes have now been identified, it is not yet known whether forced expression of any of these factors can reprogram non-lymphoid cells to become lymphocytes. These putative regulators of lymphocyte differentiation, including Oct-2 (Muller et al., 1988; Scheidereit et al., 1988; Staudt et al., 1988), PU.1 (Klemsz et al., 1990), EBF (Hagman et al., 1991), BSAP (Barberis et al., 1990), LEF-1 (Travis et al., 1991; Waterman et al., 1991), TCF-1 (van de Wetering et al., 1991), and GATA-3 (Ho et al., 1991; Joulin et al., 1991), differ in their expression patterns within the lymphocyte lineage. Oct-2 and PU.1 are primarily B cell-specific, EBF and BSAP are early B

cell-specific, LEF-1 is pre-B and T cell-specific, and TCF-1 and GATA-3 are T cell-specific. They also differ in the target genes and DNA sequence elements they recognize, and most of them have distinct DNA binding domain types. This group of putative regulators of the lymphocyte lineage is considerably more eclectic than the group of regulators of the muscle cell lineage. All of the known regulators of the muscle cell lineage (MyoD, Myogenin, Myf-5 and MRF4), are closely related basic helix-loop-helix domain proteins which have very similar or identical expression patterns (skeletal muscle cells and their precursors) and all appear to operate by recognizing the same element in the regulatory sequences of their muscle-specific target genes (Weintraub et al., 1991a).

The greater diversity of the known regulators of the lymphocyte lineage compared with those of the muscle cell lineage may reflect a more elaborate process of differentiation for lymphocytes than for muscle cells. Two properties of lymphocyte differentiation are consistent with this notion. First, lymphocyte differentiation occurs in stages in which lineage decisions are made, specific gene rearrangements occur, and sets of genes are turned on and off, all in a temporally ordered fashion. It is unlikely that a single regulatory event (analogous to the activation of the MyoD family of regulators) could account for this complex, temporally regulated process. Second, lymphocyte differentiation may not require a single, discrete switch since it is not tightly coupled to pattern formation, the way muscle differentiation must be. By contrast, a discrete nodal point (the *myoD* family) in the specification of the muscle cell lineage (Weintraub et al., 1991a), may be essential for proper muscle formation since the pattern forming constraints (the location and morphology of muscle tissue) may require that a single switch exist to allow the entire muscle cell program to be activated at once in response to the appropriate temporal and spatial signals during embryogenesis. From this perspective, the lymphoid-specific factors (including Oct-2, PU.1, EBF, BSAP,

LEF-1, TCF-1, and GATA-3), would appear to be regulators of a more complex differentiation process that occurs in multiple steps, rather than a single step.

The question arises as to how the expression of multiple lymphocyte-specific regulators can be coordinated. By analogy with the *myoD* family of regulators, the regulators of lymphocyte differentiation may form an interactive network by regulating each other. However, this putative cross-regulation would be considerably more complex than that occuring between the members of the *myoD* family, where cross-activation may simply ensure that all the members are coordinately activated (Weintraub, 1991a). Instead, interactions between the regulators of lymphocyte differentiation could involve both cross-activation and cross-repression in a regulatory cascade, analogous to the interactions that occur between regulators of early *Drosophila* embryogenesis (DiNardo et al., 1988; reviewed in Davidson, 1990). Of the putative regulators of lymphocyte differentiation that have been identified so far, LEF-1 is the one most likely to occupy an early position on such a regulatory cascade, since it is expressed in both pre-B and T lymphocytes and may also be expressed in the lymphoid stem cells that give rise to both B and T lymphocytes.

Future directions

A variety of biological and biochemical questions concerning the putative regulators of lymphoid-specific gene expression isolated in this work remain unanswered. Future lines of inquiry concerning the biology of LEF-1 include defining the function it serves in lymphocyte differentiation by disrupting the gene in the mouse germ line, as well as determining if it can initiate the lymphocyte lineage program when it is expressed artificially in non-lymphoid cells. Biochemical and biophysical pursuits include determining the structure of the HMG domain and how it binds DNA. In addition, it is now known that the binding of LEF-1

introduces a 130^{o} bend in DNA (K. Giese, J. Cox, and R. Grosschedl, unpublished results). The mechanism and function of bending are not yet known, so greater understanding of this feature will be informative. Finally, defining how the expression of the LEF-1 gene is regulated in a pre-B and T cell-specific manner may eventually make it possible to identify the molecules that specify the lymphocyte lineage.

Questions concerning the B cell-specific FB1 cDNA clone include: does the encoded polypeptide bind DNA in a sequence-specific manner, does the amino acid sequence reveal a known DNA binding domain, can putative B cell-specific target genes containing binding sites for the FB1-encoded protein be identified, and if so, is it a transcriptional activator of these genes?

The purification of EBF and recent cloning of a putative EBF cDNA (J. Hagman, A. Travis, and R. Grosschedl, unpublished data), potentially opens several directions of inquiry. Future directions concerning the biology of EBF include defining its role in the context-dependent transcriptional activation of the *mb-1* gene, determining if forced expression of EBF can activate the *mb-1* gene in cells that do not normally express it, and determining if it has a more general role in early B lymphocyte differentiation by observing the effect of disrupting the gene in the mouse germ line. In addition, by mapping its chromosomal location, it may be possible to correlate the EBF gene with a previously identified chromosomal breakpoint mutation effecting lymphocyte growth or differentiation. From a biochemical perspective, it will be informative to determine its mechanism of DNA binding and if an interaction between EBF and another DNA binding protein is involved in the context-dependent function of the EBF site in the *mb-1* promoter. Finally, defining how the EBF gene itself is regulated in an early B cell-specific manner may provide further insights into the regulatory mechanisms that direct B lymphocyte differentiation, and cell lineage specification in general.

REFERENCES

Abel, T. and T. Maniatis. 1989. Action of leucine zippers. *Nature* **341**: 24-25.

Altaba, A.R., H. Perry-O'Keefe, and D.A. Melton. 1987. *Xfin:* an embryonic gene encoding a multifingered protein in *Xenopus*. *EMBO J.* **6**: 3065-3070.

Banerji, J., L. Olson, and W. Schaffner. 1983. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell* **33**: 729-740.

Banerji, J., S. Rusconi, and W. Schaffner. 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**: 299-308.

Barberis, A., K. Widenhorn, L. Vitelli, and M. Busslinger. 1990. A novel B-cell lineage-specific transcription factor present at early but not late stages of differentiation. *Genes Dev.* **4**: 849-859.

Baumhueter, S., D.B. Mendel, P.B. Conley, C.J. Kuo, C. Turk, M.K. Graves, C.A. Edwards, G. Courtois, and G.R. Crabtree. 1990. HNF-1 shares three sequence motifs with the POU domain proteins and is identical to LF-B1 and APF. *Genes Dev.* **4**: 372-379.

Benezra, R., R.L. Davis, D. Lockshon, D.L. Turner, and H. Weintraub. 1990. The protein Id: A negative regulator of helix-loop-helix DNA binding proteins. *Cell* **61**: 49-59.

Ben-Neriah, Y., A. Bernards, M. Paskind, G.Q. Daley, and D. Baltimore. 1986. Alternative 5' exons in *c-abl* mRNA. *Cell* **44**: 577-586.

Birkmann, A. and A. Bock. 1989. Characterization of a *cis* regulatory DNA element necessary for formate induction of the formate dehydrogenase gene (fdhF) of *Escherichia coli*. *Mol. Microbiol.* **3**: 187-195.

Bodner, M., J.-L. Castrillo, L.E. Theill, T. Deerinck, M. Ellisman, and M. Karin. 1988. The pituitary-specific transcription factor GHF-1 is a homeobox-containing protein. *Cell* **55**: 505-518.

Blackman, M., J. Kappler, and P. Marrack. 1990. The role of the T cell receptor in positive and negative selection of develoing T cells. *Science* **248**: 1335-1341.

Blackwell, T.K. and F.W. Alt. 1988. Immunoglobulin genes. In *Molecular immunology* (ed. B.D. Hames and D.M. Glover), pp.1-60. IRL Press, Washington, D.C.

Blackwell, T.K. and H. Weintraub. 1990. Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science* **250**: 1104-1110.

Blackwell, T.K., L. Kretzner, E.M. Blackwood, R.N. Eisenman, and H. Weintraub. 1990. Sequence-specific DNA binding by the c-Myc protein. *Science* **250**: 1149-1151.

Blackwood, E. M. and R.N. Eisenman. 1991. Max: A helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. *Science* **251**: 1211-1217.

Braun, T., G. Buschhausen-Denker, E. Bober, E. Tannich, and H.H. Arnold. 1989. A novel human muscle factor related to but distinct from MyoD1 induces myogenic conversion in 10T1/2 fibroblasts. *EMBO J.* **8**: 701-709.

Braun, T., E. Boder, B. Winter, N. Rosenthal, and H.H. Arnold. 1990. *Myf-6*, a new member of the human gene family of myogenic determination factors: evidence for a gene cluster on chromosome 12. *EMBO J.* **9**: 821-831.

Breathnach, R. and P. Chambon. 1981. Organization and expression of eucaryotic split genes coding for proteins. *Ann. Rev. Biochem.* **50**: 349-383.

Buck, M., S. Miller, M. Drummond, and R. Dixon. 1986. Upstream activator sequences are present in the promoters of nitrogen fixation genes. *Nature* **320**: 374-378.

Buratowski, S., S. Hahn, L. Guarente, and P.A. Sharp. 1989. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56**: 549-561.

Caudy, M., E.H. Grell, C. Dambly-Chaudiere, A. Ghysen, L.Y. Jan, and Y.N. Jan. 1988. The maternal sex determination gene *daughterless* has zygotic activity necessary for the formation of peripheral neurons in *Drosophila*. *Genes Dev.* **2**: 843-852.

Cherayil, B.J. and S. Pillai. 1991. The ω/λ_5 surrogate immunooglobulin light chain is expressed on the surface of transitional B lymphocytes in murine bone marrow. *J. Exp. Med.* **173**: 111-116.

Choi, O.-R. and J.D. Engel. 1986. A 3' enhancer is required for temporal and tissue-specific transcriptional activation of the chicken adult β-globin gene. *Nature* **323**: 731-734.

Cline, T.W. 1989. The affairs of *daughterless* and the promiscuity of developmental regulators. *Cell* **59**: 231-234.

Concino, M. F., R. F. Lee, J. P. Merryweather, and R. Weinmann. 1984. The adenovirus major late promoter TATA box and initiation site are both necessary for transcription in vitro. *Nucleic Acid Res.* **12**: 7423-7433.

Cooper, M.D. and P.D. Burrows. 1989. B-cell differentiation. In *Immunoglobulin genes* (ed. T. Honjo, F.W. Alt, and T.H. Rabbits), pp.1-22. Academic Press, London.

Curran, T. and B.R. Franza Jr. 1988. Fos and Jun: the AP-1 connection. *Cell* **55**: 395-397.

Davidson, E.H. 1990. How embryos work: a comparative view of diverse modes of cell fate specification. *Development* **108**: 365-389.

Davis, M.M. and P.J. Bjorkman. 1988. T-cell antigen receptor genes and T-cell recognition. *Nature* **334**: 395-401.

Davis, R.L., H. Weintraub, and A.B. Lassar. 1987. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**: 987-1000.

Davis, R.L., P.-F. Cheng, A.B. Lassar, and H. Weintraub. 1990. The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. *Cell* **60**: 733-746.

Davison, B.L., J.-M. Egly, E.R. Mulvihill, and P. Chambon. 1983. Formation of stable preinitiation complexes between eukaryotic class B transcription factors and promoter sequences. *Nature* **301**: 680-686.

Descombes, P., M. Chojkier, S. Lichtsteiner, E. Falvey, and U. Schibler. 1990. LAP, a novel member of the C/EBP gene family, encodes a liver-enriched transcriptional activator protein. *Genes Dev.* **4**: 1541-1551.

Diamond, M.I., J.N. Miner, S.K. Yoshinaga, and K.R. Yamamoto. 1990. Transcription factor interactions: Selectors of positive or negative regulation from a single DNA element. *Science* **249**: 1266-1272.

Dignam, J.D., R.M. Lebovitz, and R.G. Roeder. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucl. Acids. Res.* **11**: 1475-1489.

DiNardo, S., E. Sher, J. Heemskerk-Jongens, J.A. Kassis, and P.H. O'Farrell. 1988. Two-tiered regulation of spatially patterned *engrailed* gene expression during *Drosophila* embryogenesis. *Nature* **332**: 604-609.

Dynlacht, B.D., T. Hoey, and R. Tjian. 1991. Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. *Cell* **66**: 563-576.

Edmondson, D.G. and E.N. Olson. 1989. A gene with homology to the *myc* similarity region of MyoD1 is expressed during myogenesis and is sufficient to activate the muscle differentiation program. *Genes Dev.* **3**: 628-640.

Ellis, H.M., D.R. Spann, and J.W. Posakony. 1990. *extramacrochaetae*, a negative regulator of sensory organ development in Drosophila, defines a new class of helix-loop-helix proteins. *Cell* **61**: 27-38.

Evans, T. and G. Felsenfeld. 1989. The erythroid-specific transcription factor Eryf1: a new finger protein. *Cell* **58**: 877-885.

Frain, M., G. Swart, P. Monaci, A. Nicosia, S. Stampfli, R. Frank, and R. Cortese. 1989. The liver-specific transcription factor LF-B1 contains a highly diverged homeobox DNA binding domain. *Cell* **59**: 145-157.

Freedman, L.P., B.F. Luisi, Z.R. Korszun, R. Basavappa, P.B. Sigler, and K.R. Yamamoto. 1988. The function and structure of the metal coordination sites within the glucocorticoid receptor DNA binding domain. *Nature* **334**: 543-546.

Garrell, J. and J. Modolell. 1990. The Drosophila *extramacrochaetae* locus, an antagonist of proneural genes that, like these genes, encodes a helix-loop-helix protein. *Cell* **61**: 39-48.

Ghosh, S., A.M. Gifford, L.R. Riviere, P. Tempst, G.P. Nolan, and D. Baltimore. 1990. Cloning of the p50 DNA binding subunit of NF-κB: homology to *rel* and *dorsal*. *Cell* **62**: 1019-1029.

Ghysen A. and C. Dambly-Chaudiere. 1989. Genesis of the Drosophila peripheral nervous system. *Trends Genet.* **4**: 251

Giese, K., A. Amsterdam, and R. Grosschedl. 1991. DNA-binding properties of the HMG domain of the lymphoid-specific transcriptional regulator LEF-1. *Genes Dev.* **5** (in press).

Gillies, S.D., S.L. Morrison, V.T Oi, and S. Tonegawa. 1983. A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell* **33**: 717-728.

Gralla, J.D. 1991. Transcriptional control-lessons from an E. coli promoter data base. *Cell* **66**: 415-418.

Greenblatt, J. 1991. Roles of TFIID in transcriptional initiation by RNA polymerase II. *Cell* **66**: 1067-1070.

Gregor, P.D., M. Sawadogo, and R.G. Roeder. 1990. The adenovirus major late transcription factor USF is a member of the helix-loop-helix group of regulatory proteins and binds to DNA as a dimer. *Genes Dev.* **4**: 1730-1740.

Grosschedl, R. and M.L. Birnstiel. 1980. Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants in vivo. *Proc. Natl. Acad. Sci. USA* **77**: 1432-1436.

Grosveld, G.C., C.K. Shewmaker, P. Jat, and R.A. Flavell. 1981. Localization of DNA sequences necessary for transcription of the rabbit β-globin gene in vitro. *Cell* **25**: 215-226.

Guazzi, S., M. Price, M De Felice, G. Damante, M.-G. Mattei, and R Di Lauro. 1990. Thyroid nuclear factor 1 (TTF-1) contains a homeodomain and displays a novel DNA binding specificity. *EMBO J.* **9**: 3631-3639.

Gubbay, J., J. Collignon, P. Koopman, B. Capel, A. Economou, A. Munsterberg, N. Vivian, P, Goodfellow, and R. Lovell-Badge. 1990. A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature* **346**: 245-250.

Gunther, C.V., J.A. Nye, R.S. Bryner, and B.J. Graves. 1990. Sequence-specific DNA binding of the proto-oncoprotein *ets-1* defines a transcriptional activator sequence within the long terminal repeat of the Moloney murine sarcoma virus. *Genes Dev.* **4**: 667-679.

Hager, D.A. and R.R. Burgess. 1980. Elution of protein from sodium dodecyl sulfate-polyacrylamide gels, removal of sodium dodecyl sulfate, and renaturation of enzymatic activity: results with sigma subunit of *E. coli* RNA polymerase, wheat germ DNA topoisomerase, and other enzymes. *Anal. Biochem.* **109**: 76-86.

Hagman, J., A. Travis, and R. Grosschedl. 1991. A novel lineage-specific nuclear factor regulates mb-1 gene transcription at the early stages of B cell differentiation. *EMBO J.* **10**: 3409-3417.

Harlow, E. and D. Lane. 1988. *Antibodies: A laboratory manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Hermanson, G.G., M. Briskin, D. Sigman, and R. Wall. 1989. Immunoglobulin enhancer and promoter motifs 5' of the B29 B-cell-specific gene. *Proc. Natl. Acad. Sci. USA* **86**: 7341-7345.

Hermanson, G.G., D. Eisenberg, P.W. Kincade and R. Wall. 1988. B29: A member of the immunoglobulin gene superfamily exclusively expressed on B-lineage cells. *Proc. Natl. Acad. Sci. USA* **85**: 6890-6894.

Herr, W., R.A. Sturm, R.G. Clerc, L.M. Corcoran, D. Baltimore, P.A. Sharp, H.A. Ingraham, M.G. Rosenfeld, M. Finney, G. Ruvkun, and H.R. Horvitz. 1988. The POU domain: A large concerved region in the mmammalian *pit-1*, *oct-1*, *oct-2*. and *Caenorhabiditis elegans unc-86* gene products. *Genes and Dev.* **2**: 1513-1516.

Hesse, J.E, J.M. Nickol, M.R. Lieber, and G. Felsenfeld. 1986. Regulated gene expression in transfected primary chicken erythrocytes. *Proc. Natl. Acad. Sci. USA* **83**: 4312-4316.

Ho, I.-C., L.-H. Yang, G. Morle, and J.M. Leiden. 1989. A T-cell-specific transcriptional enhancer element 3' of the Cα in the human T-cell receptor α locus. *Proc. Natl. Acad. Sci. USA* **86**: 6714-6718.

Ho, I.-C., P. Vorhees, N. Marin, B. Karpinski Oakley, S.-F. Tsai, S. H. Orkin, and J.M. Leiden. 1991. Human GATA-3: a lineage-restricted transcription factor that regulates the expression of the T cell receptor a gene. *EMBO J.* **10**: 1187-1192.

Hoover, T.R., E. Santero, S. Porter, and S. Kustu. 1990. The integration host factor stimulates interaction of RNA polymerase with NIFA, the transcriptional activator for nitrogen fixation operons. *Cell* **63**: 11-22.

Hope, I.A. and K. Struhl. 1986. Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. *Cell* **46**: 885-894.

Hopwood, N.D., A. Pluck, and J.B. Gurdon. 1989. MyoD expression in the forming somites is an early response to mesoderm induction in Xenopus embryos. *EMBO J.* **8**: 3409-3417.

Hu, Y.-F., B. Luscher, A. Admon, N. Mermod, and R. Tjian. 1990. Transcription factor AP-4 contains multiple dimerization domains that regulate dimer specificity. *Genes Dev.* **4**: 1741-1752.

Ingraham, H.A., R. Chen, H.J. Mangalam, H.P. Elsholtz, S.E. Flynn, C.R. Lin, D.M. Simmons, L. Swanson, and M.G. Rosenfeld. 1988. A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype. *Cell* **55**: 519-529.

Jacob, F. and J. Monod. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**: 318-356.

Jantzen, H.-M., A. Admon, P.B. Bell, and R. Tjian. 1990. Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins. *Nature* **344**: 830-836.

Jones, N. 1990. Transcriptional regulation by dimerization: two sides to an incestuoius relationship. *Cell* **61**: 9-11.

Joulin, V., D. Bories, J.-F. Eleouet, M.-C. Labastie, S. Chretien, M.-G. Mattei, and P.-H. Romeo. 1991. A T-cell specific TCR δ DNA binding protein is a member of the human GATA family. *EMBO J.* **10**: 1809-1816.

Kadonaga, J.T., K.R. Carner, F.R. Masiarz, and R. Tjian. 1987. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. *Cell* **51**: 1079-1090.

Kadonaga, J.T., and R. Tjian. 1986. Affinity purification of sequence-specific DNA binding proteins. *Proc. Natl. Acad. Sci. USA* **83**: 5889-5893.

Karlsson, O., S. Thor, T. Norberg, H. Ohlsson, and T. Edlund. 1990. Insulin gene enhancer binding protein Isl-1 is a member of a novel class of proteins containing both a homeo- and a Cys-His domain. *Nature* **344**: 879-882.

Kelly, M., J. Burke, M. Smith, A. Klar, and D. Beach. 1988. Four mating-type genes control sexual differentiation in the fission yeast. *EMBO J.* **7**: 1537-1547.

Kessel, M. and P. Gruss. 1990. Murine developmental control genes. *Science* **249**: 374-379.

Kissinger, C.R., B. Liu, E. Martin-Blanco, T.B. Kornberg, and C.O. Pabo. 1990. Crystal structure of an engrailed homeodomain-DNA complex at 2.8 A resolution: A framework for understanding homeodomain-DNA interactions. *Cell* **63**: 579-590.

Klemsz, M.J., S.R. McKercher, A. Celada, C.Van Beveren, and R.A. Maki. 1990. The macrophage and B cell-specific transcription factor PU.1 is related to the *ets* oncogene. *Cell* **61**: 113-124.

Kudo, A., and F. Melchers. 1987. A second gene, V_{preB} in the λ_5 locus of the mouse, which appears to be selectively expressed in pre-B lymphocytes. *EMBO J.* **6**:2267-2272.

Lai, E., V.R. Prezioso, E. Smith, O. Litvin, R.H. Costa, and J.E. Darnell, Jr. 1990. HNF-3A, a hepatocyte-enriched transcription factor of novel structure is regulated transcriptionally. *Genes Dev.* **4**: 1427-1436.

Lai, E., V.R. Prezioso, W. Tao, W.S. Chen, and J.E. Darnell, Jr. 1991. Hepatocyte nuclear factor 3α belongs to a gene family in mammals that is homologous to the *Drosophila* homeotic gene *fork head*. *Genes Dev.* **5**: 416-427.

Landau, N.R., D.G. Schatz, M. Rosa, and D. Baltimore. 1987. Increased frequency of N-region insertion in a murine pre-B-cell line infected with a terminal deoxynucleotidyl transferase retroviral expression vector. *Mol. Cell. Biol.* **7**: 3237-3243.

Landau, N.R., T.P. St. John, I.L. Weissman, S.C. Wolf, A.E. Silverstone, and D. Baltimore. 1984. Cloning of terminal transferase cDNA by antibody screening. *Proc. Natl. Acad. Sci. USA* **81**: 5836-5840.

Landschultz, W.H., P.F. Johnson, S.L. McKnight. 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240**: 1759-1764.

Lassar, A.B., R.L. Davis, W.E. Wright, T. Kadesch, C. Murre, A. Voronova, D. Baltimore, and H. Weintraub. 1991. Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo. *Cell* **66**: 305-315.

Lassar, A.B., B.M. Paterson, and H. Weintraub. 1986. Transfection of a DNA locus that mediates the conversion of 10T1/2 fibroblasts to myoblasts. *Cell* **47**: 649-656.

Lewis, M.K. and R.R. Burgess. 1982. Eukaryotic RNA polymerases. In *The enzymes, vol. XV.*, pp.109-153. Academic Press, Inc.

Lillie, J.W., and M.R. Green. 1989. Activator's target in sight. *Nature* **341**: 279-280.

Lin, Y.-S. and M.R. Green. 1991. Mechanism of action of an acidic transcriptional activator in vitro. *Cell* **64**: 971-981.

Lin, Z., C.A. Dechesne, J. Eldridge, and B.M. Paterson. 1989. An avian muscle factor related to MyoD1 activates muscle-specific promoters in nonmuscle cells of different germ-layer origin and in BrdU treated myoblasts. *Genes Dev.* **3**: 986-996.

Lo, K., N.R. Landau, and S.T. Smale. 1991. LyF-1, a transcriptional regulator that interacts with a novel class of promoters for lymphocyte-specific genes. *Mol. Cel. Biol.* **11**: 5229-4243.

Luisi, B.F., W.X. Xu, Z. Otwinowski, L.P. Freedman, K.R. Yamamoto, and P.B. Sigler. 1991. Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* **352**: 497-505.

Luscher, B. and R.N. Eisenman. 1990. New light on Myc and Myb. Part I. Myc. *Genes Dev.* **4**: 2025-2035.

Magasanik, B. 1989. Gene regulation from sites near and far. *The New Biol.* **1**: 247-251.

McDougall, S., C.L. Peterson, and K. Calame. 1988. A transcriptional enhancer 3' of the Cβ2 in the T cell receptor β locus. *Science* **241**: 205-208.

McGinnis, W., M.S. Levine, E. Hafen, A. Kuroiwa, and W.J. Gehring. 1984. A conserved DNA sequence in homeotic genes of the Drosophila Antennapedia and Bithorax complexes. *Nature* **308**: 428-433.

McKnight, S. L., and R. Kingsbury. 1982. Transcriptional control signals of a eukaryotic protein-coding gene. *Science* **217**: 316-324.

Meisterernst, M. and R.G. Roeder. 1991. Family of proteins that interact with TFIID and regulate promoter activity. *Cell* **67**: 557-567.

Mendel, D.B., L.P. Hansen, M.K. Graves, P.B. Conley, and G.R. Crabtree. 1991. HNF-1α and HNF-1β (vHNF-1) share dimerization and homeo domains, but not activation domains, and form heterodimers in vitro. *Genes Dev.* **5**: 1042-1056.

Miesfeld, R., S. Rusconi, P.J. Godowski, B.A. Maler, S. Okret, A.-C. Wilkstrom, J.-A. Gustafsson, K.R. Yamamoto. 1986. Genetic Complementation of a glucocorticoid receptor deficiency by expression of cloned receptor cDNA. *Cell* **46**: 389-399.

Miller, J., A.D. McLachlan, and A. Klug. 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO J.* **4**: 1609-1614.

Mitchell, P.J. and R. Tjian. 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245**: 371-378.

Moreau, P., R. Hen, B. Wasylyk, R. Everett, M.P. Graub, and P. Chambon. 1981. The SV40 72 base pair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucl. Acids Res.* **9**: 6047-6068.

Mueller, C.R., P. Maire, and U. Schibler. 1990. DBP, a liver-enriched transcriptional activator, is expressed late in ontogeny and its tissue specificity is determined postranscriptionally. *Cell* **61**: 279-291.

Mueller, P.R. and B. Wold. 1989. *In vivo* footprinting of a muscle specific enhancer by ligation mediated PCR. *Science* **246**: 780-786.

Muller, H.-P., J.M. Sogo, and W. Schaffner. 1989. An enhancer stimulates transcription in *trans* when attached to the promoter via a protein bridge. *Cell* **58**: 767-777.

Muller, M.M., S. Ruppsert, W. Schaffner, and P. Matthias. 1988. A cloned octamer transcription factor stimulates transcription from lymphoid-spceific promoters in non-B cells. *Nature* **336**: 544-551.

Murre, C., P.S. McCaw and D. Baltimore. 1989a. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, *daughterless*, *MyoD*, and *myc* proteins. *Cell* **56**: 777-783.

Murre, C., P.S. McCaw, H. Vaessin, M. Caudy, L.Y. Jan, Y.N. Jan, C.V. Cabrera, J.N. Buskin, S.D. Hauschka, A.B. Lassar, H. Weintraub, and D. Baltimore. 1989b. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* **58**: 537-544.

Nakajima, N., M. Horikoshi, and R.G. Roeder. 1988. Factors involved in specific transcription by mammalian RNA poymerase II: purification, genetic specificity, and TATA box-promoter interactions of TFIID. *Mol. Cell. Biol.* **8**: 4028-4040.

Nelson, K., E. Mather, and R. Perry. 1984. Lipopolysaccharide-induced transcription of the κ immunoglobulin locus occurs on both alleles and is independent of methylation status. *Nucleic Acids Res.* **12**: 1911-1923.

Nolan, G.P., S. Ghosh, H.-C. Liou, P. Tempst, and D. Baltimore. 1991. DNA binding and IκB inhibition of the cloned p65 subunit of NF-κB, a *rel*-related polypeptide. *Cell* **64**: 961-969.

Oettinger, M.A., D.G. Schatz, C. Gorka, and D. Baltimore. 1990. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science* **248**: 1517-1523.

Oosterwegel, M., M. van de Wetering, D. Dooijes, L. Klomp, A. Winoto, K. Georgopoulos, F. Meijlink, and H. Clevers. 1991. Cloning of murine TCF-1, a T cell-specific transcription factor interacting with functional motifs in the CD3-ε and T cell receptor α enhancers. *J. Exp. Med.* **173**: 1133-1142.

O'Shea E.K., R.H. Rutkowski, W.S. Stafford III, P.S. Kim. 1989. Preferential heterodimer formation by isolated leucine zippers from Fos and Jun. *Science* **245**: 646-648.

Pabo, C.O. and R.T. Sauer. 1984. Protein-DNA recognition. *Annu. Rev. Biochem.* **53**: 293-321.

Parkhurst, M., D. Bopp, and D. Ish-Horowicz. 1990. X:A ratio, the primary sex-determining signal in Drosophila, is transduced by helix-loop-helix proteins. *Cell* **63**: 1179-1191.

Pavletich, N.P. and C.O. Pabo. 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science* **252**: 809-817.

Poellinger, L., B.K. yoza, and R.G. Roeder. 1989 Functional cooperativity between protein molecules bound at two distinct sequence elements of the immunoglobulin heavy chain enhancer. *Nature* **337**: 573-576.

Poole, S.J., L.M. Kauvar, B. Drees, and T. Kornberg. 1985. The *engrailed* locus of Drosophila: structural analysis of an embryonic transcript. *Cell* **40**: 37-43.

Popham, D., D. Szeto, J. Keener, and S. Kustu 1989. Function of a bacterial activator protein that binds to transcriptional enhancers. *Science* **243**: 629-635.

Prendergast, G.C., D. Lawe, and E.B. Ziff. 1991. Association of Myn, the murine homolog of Max, with c-Myc stimulates methylation-sensitive DNA binding and Ras cotransformation. *Cell* **65**: 395-407.

Ptashne, M. 1986. Gene regulation by proteins acting nearby and at a distance. *Nature* **322**: 697-701.

Ptashne, M. 1988. How eukaryotic transcriptional activators work. *Nature* **335**: 683-689.

Pugh, B.F. and R. Tjian. 1990. Mechanism of transcriptional activiation by Sp1: evidence for coactivators. *Cell* **61**: 1187-1197.

Reitzer, L.J. and B. Magasanik. 1986. Transcription of *glnA* in E. coli is stimulated by activator bound to sites far from the promoter. *Cell* **45**: 785-792.

Rhodes, S.J. and S.F. Konieczny. 1989. Identification of MRF4: a new member of the muscle regulatory factor gene family. *Genes Dev.* **3**: 2050-2061.

Romani, S., S. Campuzano, E.R. Macagno, and J. Modolell. 1989. Expression of *achaete* and *scute* genes in *Drosophila* imaginal discs and their function in senory organ development. *Genes Dev.* **3**: 997-1107.

Rosenfeld, M.G. 1991. POU-domain transcription factors: pou-er-ful developmental regulators. *Genes Dev.* **5**: 897-907.

Rupp, R.A.W. and H. Weintraub. 1991. Ubiquitous myoD transcription at the midblastula transition precedes induction-dependent myoD expression in presumptive mesoderm of X. laevis. *Cell* **65**: 927-937.

Sakaguchi, N., C. N. Berger, and F. Melchers. 1986. Isolation of a cDNA copy of an RNA species expressed in murine pre-B cells. *EMBO J.* **5**: 2139-2147.

Sakaguchi, N., S.-I. Kashiwamura, M. Kimoto, P. Thalmann, and F. Melchers. 1988. B lymphocyte lineage-restricted expression of *mb-1*, a gene with CD3-like structural properties. *EMBO J.* **7**: 3457-3464.

Saltzman, A.G., and R. Weinmann. 1989. Promoter specificity and modulation of RNA polymerase II transcription. *FASEB J.* **3**: 1723-1733.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: A laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sasse-Dwight, S. and J.D. Gralla. 1988. Probing the *Escherichia coli glnALG* upstream activation mechanism in vivo. *Proc. Natl. Acad. Sci. USA* **85**: 8934-8938.

Sawadogo, M. and R.G. Roeder. 1985. Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. *Cell* **43**: 165-175.

Schatz, D.G., M.A. Oettinger, and D. Baltimore. 1989. The V(D)J recombination activating gene, RAG-1. *Cell* **59**: 1035-1048.

Scheidereit, C., J.A. Cromlish, T. Gerster, K. Kawakami, C. Balmaceda, R.A. Currie, and R.G. Roeder. 1988. A human lymphoid-specific transcription factor that activates immunoglobulin genes is a homeo-box protein. *Nature* **336**: 551-557.

Scholer, H.R., A.K. Hatzopoulos, R. Balling, N. Suzuki, and P. Gruss. 1989. A family of octamer-specific proteins present during mouse embryogenesis: Evidence for germline-specific expression of an Oct factor. *EMBO J.* **8**: 2543-2550.

Scott, M.P., J.W. Tamkun, and G.W. Hartzell. 1989. The structure and function of the homeodomain. *Biochim. Biophys. Acta* **989**: 25-48.

Sinclair, A.H., P. Berta, M.S. Palmer, J.R. Hawkins, B.L. Griffiths, M.J. Smith, J.W. Foster, A.M. Frischauf, R. Lovell-Badge and P.N. Goodfellow. 1990. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* **346**: 240-244.

Singh, H., R. Sen, D. Baltimore, P.A. Sharp. 1986. A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. *Nature* **319**: 154-158.

Singh, H., J.H. LeBowitz, A.S. Baldwin, and P.A. Sharp. 1988. Molecular cloning of an enhancer binding protein: Isolation by screening of an expression library with a recognition site DNA. *Cell* **52**: 415-423.

Sladek, F.M., W. Zhong, E. Lai, and J.E. Darnell, Jr. 1990. Liver-enriched transcription factor HNF-4 is a novel member of the steroid hormone receptor superfamily. *Genes Dev.* **4**: 2353-2365.

Smale, S.T. and D. Baltimore. 1989. The "initiator" as a transcription control element. *Cell* **57**: 103-113.

Smale, S.T., M.C. Schmidt, A.J. Berk, and D. Baltimore. 1990. Transcriptional activation by Sp1 as directed through TATA or initiator: specific requirement for mammalian transcription factor IID. *Proc. Natl. Acad. Sci. USA* **87**: 4509-4513.

Staben, C. and C. Yanofsky. 1990. *Neurospora crassa a* mating-type region. *Proc. Natl. Acad. Sci. USA.* **87**: 4917-4921.

Staudt, L.M., R.G. Clerc, H. Singh, J.H. LeBowitz, P.A. Sharp, and D. Baltimore. 1988. Cloning of a lymphoid-specific cDNA encoding a protein binding the regulatory octamer DNA motif. *Science* **241**: 577-580.

Stringer, K.F., C.J. Ingles, and J. Greenblatt. 1990. Direct and selective binding of an acidic transcriptional activation domain to the TATA-box factor TFIID. *Nature* **345**: 783-785.

Su, W., S. Porter, S. Kustu, and H. Echols. 1990. DNA-looping and enhancer activity: association between DNA-bound NTRC activator and RNA polymerase at the bacterial *glnA* promoter. *Proc. Natl. Acad. Sci. USA* **87**: 5504-5508.

Sumimoto, H., Y. Ohkuma, T. Yamamoto, M. Horikoshi, and R.G. Roeder. 1990. Factors involved in specific transcription by mammalian RNA polymerase II: identification of general transcription factor TFIIG. *Proc. Natl. Acad. Sci. USA* **87**: 9158-9162.

Talanian, R.V., C.J. McKnight, P.S. Kim. 1990. Sequence-specific DNA binding by a short peptide dimer. *Science* **249**: 769-771.

Tedder, T. F., and C. M. Isaacs. 1989. Isolation of cDNAs encoding the CD19 antigen of human and mouse B lymphocytes. A new member of the immunoglobulin superfamily. *J. Immunol.* **143**: 712-717.

Theisen, M., A. Stief, and A.E. Sippel. 1986. The lysozyme enhancer: cell-specific activation of the chicken lysozyme gene by a far-upstream DNA element. *EMBO J.* **5**: 719-724.

Thomas, M.L. 1989. The leucocyte common antigen family. *Annu. Rev. Immunol.* **7**: 339-369.

Travis, A., A. Amsterdam, C. Belanger, and R. Grosschedl. 1991a. LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor α enhancer function. *Genes Dev.* **5**: 880-894.

Travis, A., J. Hagman, and R. Grosschedl. 1991b. Heterogeneously initiated transcription from the pre-B- and B-cell-specific *mb-1* promoter: analysis of the requirement for upstream factor-binding sites and initiation site sequences. *Mol. Cell. Biol.* **11**: 5756-5766.

Tsai, S.-F., D.I.K. Martin, L.I. Zon, A.D. D'Andrea, G.G. Wong, and S.H. Orkin. 1989. Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. *Nature* **339**: 446-451.

Vallee, B.L., J.E. Coleman, and D.S. Auld. 1991. Zinc fingers, zinc clusters, and zinc twists in DNA-binding protein domains. *Proc. Natl. Acad. Sci. USA* **88**: 999-1003.

van de Wetering, M., M. Oosterwegel, D. Dooijes and H. Clevers. 1991. Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box. *EMBO J.* **10**: 123-132.

Venkitaraman, A.R., G.T. Williams, P. Dariavach, and M.S. Neuberger. 1991. The B-cell antigen receptor of the five immunoglobulin classes. *Nature* **352**: 777-781.

Vinson, C.R., K.L. LaMarco, P.F. Johnson, W.H. Landschulz and S.L. McKnight. 1988. In situ detection of sequence-specific DNA binding activity specified by a recombinant bacteriophage. *Genes Dev.* **2**: 801-806.

Vinson, C.R., P.B. Sigler, and S. L. McKnight. 1989. Scissors-grip model for DNA recognition by a family of leucine zipper proteins. *Science* **246**: 911-916.

von Hippel, P.H., D.G. Bear, R.B. Winter, and O.G. Berg. 1982. Molecular aspects of promoter function: an overview. in: *Promoters: Structure and Function* (ed. R.L. Rodriguez and M.J. Chamberlin), pp. 3-33. Praeger Pub. Co., New York.

Voronova, A. and D. Baltimore. 1990. Mutations that disrupt DNA binding and dimer formation in the E47 helix-loop-helix protein map to distinct domains. *Proc. Natl. Acad. Sci. USA* **87**: 4722-4726.

Voss, S.D., U. Schlokat and P. Gruss. 1986. The role of enhancers in the regulation of cell-type-specific transcriptional control. *Trends Biochem. Sci.* **11**: 287-289.

Wasylyk, B., R. Derbyshire, A. Guy, D. Molko, A. Roget, R. Teoule, and P. Chambon. 1980. Specific in vitro transcription of conalbumin gene is drastically decreased by single-point mutation in TATA box homology sequence. *Proc. Natl. Acad. Sci. USA* **77**: 7024-7028.

Waterman, M.L., W.H. Fischer, and K.A. Jones. 1991. A thymus-specific member of the HMG protein family regulates the human T cell receptor Cα enhancer. *Genes Dev.* **5**: 656-669.

Watson, D.K., M.J. McWilliams, P. Lapis, J.A. Lautenberger, C.W. Schweinfest, and T.S. Papas. 1988. Mammalian *ets-1* and *ets-2* genes encode highly conserved proteins. *Proc. Natl. Acad. Sci. USA* **85**: 7862-7866.

Wedel, A., D.S. Weiss, D. Popham, P. Droge, and S. Kustu. 1990. A bacterial enhancer functions to tether a transcriptional activator near a promoter. *Science* **248**: 486-490.

Weintraub, H., S.J. Tapscott, R.L. Davis, M.J. Thayer, M.K. Adam, A.B. Lassar, and A.D. Miller. 1989. Activation of muscle specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD. *Proc. Natl. Acad. Sci. USA* **86**: 5434-5438.

Weintraub, H., R. Davis, S. Tapscott, M. Thayer, M. Krause, kR. Benezra, T.K. Blackwell, D. Turner, R. Rupp, S. Hollenberg, Y. Zhuang, and A. Lassar. 1991a. The *myoD* gene family: Nodal point during specification of the muscle cell lineage. *Science* **251**: 761-766.

Weintraub, H., V.J. Dwarki, I. Verma, R. Davis, S. Hollenberg, L. Snider, A. Lassar, and S.J. Tapscott. 1991b. Muscle-specific transcriptional activation by MyoD. *Genes Dev.* **5**: 1377-1386.

Winoto, A. and D. Baltimore. 1989. A novel, inducible and T cell-specific enhancer located at the 3' end of the T cell receptor α locus. *EMBO J.* **8**: 729-733.

Wright, W.E., D.A. Sassoon, and V.K. Lin. 1989. Myogenin, a factor regulating myogenesis, has a domain homologous to MyoD. *Cell* **56**: 607-617.

Yamamoto, M., L.J. Ko, M.W. Leonard, H. Beug, S.H. Orkin, and J.D. Engel. 1990. Activity and tissue-specific expression of the transcription factor NF-E1 multigene family. *Genes Dev* **4**: 1650-1662.

Yancopoulos, G.D. et al. 1990. Isolation of coordinately regulated genes that are expressed in discrete stages of B-cell development. *Proc. Natl. Acad. Sci. USA* **87**: 5759-5763.

Ziff, E.B. 1990. Transcription factors: a new family gathers at the cAMP response site. *Trends Genet.* **6**: 69-72.