A thymus-specific member of the HMG protein family regulates the human T cell receptor C α enhancer

Marian L. Waterman, Wolfgang H. Fischer,¹ and Katherine A. Jones²

Molecular Biology and Virology Laboratory and Regulatory Biology Laboratory, ¹Peptide Biology Laboratory, The Salk Institute, La Jolla, California 92037 USA

The human T cell-specific transcription factor TCF-1 α plays a key role in the tissue-specific activation of the T cell receptor (TCR) C α enhancer and binds to pyrimidine-rich elements (5'-PyCTTTG-3') present in a variety of other T cell-specific control regions. Using amino acid sequence information derived from the DNA affinity-purified protein, we have now isolated cDNA clones encoding TCF-1 α . The TCF-1 α cDNA contains a single 68-amino-acid domain that is homologous to a region conserved among high-mobility group (HMG) and nonhistone chromosomal proteins. Expression of full-length and mutant cDNA clones in bacteria reveal that the single HMG motif, which is predicted to contain two extended α -helical segments, is sufficient to direct the sequence-specific binding of TCF-1 α to DNA. Northern blot experiments demonstrate further that TCF-1 α mRNA is highly tissue specific, found primarily in the thymus or T cell lines. The immature CEM T cell line expresses relatively low levels of TCF-1 α mRNA, which are increased upon activation of these cells by phorbol esters. Interestingly, the cloned TCF-1 α protein is a potent transcriptional activator of the human TCR α enhancer in nonlymphoid cell lines, whereas the activity of the endogenous protein in T cell lines is strongly dependent on an additional T cell-specific protein that interacts with the core enhancer. TCF-1 α is currently unique among the newly emerging family of DNA-binding regulatory proteins that share the HMG motif in that it is a highly tissue-specific RNA polymerase II transcription factor.

[Key Words: TCR α enhancer; HMG protein; T cell-specific transcription factor]

Received January 2, 1991; revised version accepted January 25, 1991.

The development of pluripotent stem cells into highly specialized T cell lymphocytes can be defined by the appearance and assembly at the cell surface of multicomponent complexes containing the T cell receptor (TCR) and CD3 antigens. Functionally distinct T cell lineages bear either the α/β or γ/δ heterodimeric forms of the T cell receptor (Davis and Bjorkman 1988). TCR α/β ⁺ cells predominate in mature peripheral blood T cells (including helper and cytotoxic cell subtypes) and function to mediate major histocompatibility complex (MHC)-restricted antigen recognition in combination with either the CD4 or CD8 antigens (for review, see Marrack and Kappler 1987; Strominger 1989). The TCR genes, like other members of the immunoglobulin gene superfamily, are rearranged and expressed developmentally in a stage-specific manner that begins with the γ/δ -chain genes and follows with the β -chain and lastly the α -chain genes (Furley et al. 1986; Pardoll et al. 1987; Havran and Allison 1988). The coordinate synthesis of the members of the TCR-CD3 complex is accomplished by a combination of both transcriptional and post-transcriptional steps (Wilkinson and MacLeod 1988; Paillard et al. 1990),

with transcriptional regulation mediated by potent T cell-specific enhancers that have been mapped downstream of many of these genes (Krimpenfort et al. 1988; McDougall et al. 1988; Ho et al. 1989; Winoto and Baltimore 1989a; Redondo et al. 1990).

The best-characterized of the TCR enhancers lies 4.5 kb downstream of the α -chain gene (Chien et al. 1987) and is recognized by a complex array of DNA-binding proteins, although a minimal core segment spanning 100 bp is sufficient for high-level activity in mature α/β ⁺ T cell lines (Ho et al. 1989; Winoto and Baltimore 1989a). This core segment interacts with three DNA-binding proteins: a ubiquitously distributed member of the cAMP-responsive (CREB/ATF) transcription factor family (Ho et al. 1989), the T cell-specific TCF-1 α factor (Waterman and Jones 1990), and a distinct T cell-specific protein (TCF-2 α) that may be a member of the *c-ets* protein family (Ho et al. 1990). The core TCR α enhancer is a potent T cell-specific control region that can activate heterologous promoters in mature T cell lines, and each of the three protein-binding sites is necessary for enhancer activity in vivo (Winoto and Baltimore 1989a; Ho et al. 1990; Waterman and Jones 1990). Interestingly, a subfragment of the enhancer lacking the CRE motif re-

²Corresponding author.

presses transcription from heterologous promoters in a T cell-specific manner (Ho and Leiden 1990). The α enhancer is less active in immature T cell lines (e.g., CCRF-CEM cells) but can be stimulated in these cells by treatment with phorbol esters (Winoto and Baltimore 1989a), which also induces surface expression of the TCR α/β complex (Shackelford et al. 1987).

Rearrangement and expression of the TCR α gene is regulated in a lineage-specific manner in developing thymocytes because the α -chain gene is transcribed in TCR α/β^+ but not in TCR γ/δ^+ T cells (for review, see Alt et al. 1987). In the mouse, the minimal 116-bp TCR α enhancer core was found to be fully active in γ/δ^+ cell lines, whereas larger fragments of the enhancer (5 kb or greater) were inactive (Winoto and Baltimore 1989b). These larger fragments contain multiple "silencer" elements reported to extinguish enhancer activity in a variety of non- $\alpha\beta^+$ cell lines. Consequently, the T cell-specific factors associated with the minimal enhancer appear to be expressed and active in both T cell subtypes but are subject to lineage-specific repression in γ/δ^+ T cells.

We recently reported the purification and characterization of TCF-1 α , a new T cell-specific DNA-binding protein that recognizes the 100-bp core region of the human TCR α enhancer (Waterman and Jones 1990). TCF-1 α was purified from nuclear extracts of Jurkat cells and shown to consist of a family of 57- to 53-kD proteins that bind specifically to pyrimidine-rich elements (e.g., 5'-PyCTTTG-3'), which are present in a variety of T cell-specific control regions, including the TCR α enhancer, the p56^{lck} and CD3 (γ and δ) promoters, and the HIV-1 long terminal repeat (LTR). More recently, we have found that the TCF-1 α protein also interacts with the TCR δ enhancer (M.L. Waterman and K. Jones, unpubl.), as well as with transcriptional regulatory regions of the murine CD4 gene (M.L. Waterman, K. Jones, J. Siu, and S. Hedrick, unpubl.). A double point mutation within the TCR α enhancer (5'-PyCTTTG-3' to 5'-PyCATAG-3') that eliminates the binding of affinity-purified TCF-1 α was found to decrease TCR α enhancer activity ~100-fold in vivo (Waterman and Jones 1990). This mutation also destroyed the striking synergism observed between tandem copies of the TCR α enhancer, effectively eliminating T cell-specific transcriptional activity in vivo, and reduced enhancer activity in an in vitro transcription system derived from Jurkat nuclei. The strong activation provided by TCF-1 α was highly dependent on its context with the TCR α enhancer, because constructs bearing up to four tandem copies of the isolated TCF-1 α -binding site did not demonstrate appreciable T cell-specific enhancer activity in vivo or in vitro in the absence of the flanking CRE and TCF-2 α /*ets*-binding site (Waterman and Jones 1990). These observations raise the interesting possibility that TCF-1 α interacts directly with the CREB and TCF-2 α /*ets* proteins to activate the TCR α enhancer.

To better define the role of this transcription factor in T lymphocytes, we have isolated and characterized cDNA clones encoding TCF-1 α . Most strikingly, the

TCF-1 α transcription factor was found to harbor a single 68-amino-acid domain that is homologous to a conserved region of the high-mobility group (HMG) and other nonhistone chromosomal proteins. This HMG motif is shown to be both necessary and sufficient for specific binding of TCF-1 α to its conserved pyrimidine-rich DNA element. The cloned TCF-1 α protein can also strongly activate the expression of a minimal promoter linked to the TCR α enhancer in nonlymphoid cells. Expression of TCF-1 α mRNA is highly restricted to the thymus and to T cell lines and can be further induced upon activation of an immature T cell line. These data establish that TCF-1 α belongs to a new family of regulatory proteins related by the shared HMG box homology. Other members of this family include hUBF, a ubiquitous RNA polymerase I enhancer-binding protein (Jantzen et al. 1990), and SRY, a testis-specific protein encoded on the Y chromosome that is implicated genetically in human sex determination (Gubbay et al. 1990; Sinclair et al. 1990). TCF-1 α is presently unique among this family in that its specific DNA-binding activity is determined by a single HMG motif, and it is required for the T cell-specific regulation of RNA polymerase II genes.

Results

Amino acid sequence analysis of DNA affinity-purified TCF-1 α

Our previous analysis of the T cell-specific TCF-1 α factor indicated that it was a unique protein that had not been implicated previously in T cell-specific transcriptional regulation (Waterman and Jones 1990). To isolate cDNAs encoding TCF-1 α and further characterize this regulatory protein, ~200 pmoles (16 μ g) of TCF-1 α was purified from Jurkat nuclear extracts, using a combination of conventional gradient and DNA affinity chromatography. TCF-1 α activity derives from a family of 57- to 53-kD proteins, each of which is capable of binding DNA specifically in Southwestern blotting as well as gel excision and renaturation experiments (Waterman and Jones 1990). Three distinct TCF-1 α proteins copurified with a nonspecific DNA-binding protein of 116 kD during the purification procedure. The proteins were resolved on a preparative denaturing SDS-polyacrylamide gel, transferred electrophoretically to an Immobilon filter, and visualized with Amido Black stain (Fig. 1A). The major TCF-1 α species (a single band of 55 kD) was eluted from the Immobilon membrane and digested with trypsin (Fischer et al. 1991). Individual tryptic peptides were then purified by HPLC chromatography for microsequencing (Fig. 1B), and partial amino acid sequence information was obtained from six distinct peptides of the purified TCF-1 α protein (Table 1).

Sequence information from two of the peptides (G537 and G538) was used to design combined degenerate and "guess-mer" oligodeoxynucleotides to different regions of the TCF-1 α gene, following the guidelines established by Lathe (1985). These DNA probes (Table 1) were an-

Figure 1. Purification of TCF-1 α for amino acid sequence analysis. (A) A total of 16 μ g of purified Jurkat TCF-1 α cells was transferred to an Immobilon membrane following 10% SDS-PAGE, and transferred proteins were visualized by Amido Black staining. Brackets indicate the 57- to 53-kD TCF-1 α proteins; a 116-kD nonspecific DNA-binding protein contaminant is also marked. (→) The 55-kD band excised for tryptic digestion and amino acid sequencing. (B) Reverse-phase separation of TCF-1 α tryptic peptides on a Vydac C-18 HPLC column (dimensions: 2.1 $\times$ 150 mm; pore size, 300 Å; particle size, 5 μ m). G537–G541 refer to six peaks that were collected manually and used for amino acid sequence determination. (C) Schematic diagram of the 3.2-kb TCF-1 α cDNA clone. Open boxes represent 5' and 3' UTRs, and the ORF is indicated with a line. Black boxes show the relative location of the six sequenced tryptic fragments. The hatched bar indicates the DNA fragment obtained by PCR amplification, using oligonucleotides complementary to peptides G538 and G537 (arrows).

nealed to cDNA prepared from Jurkat poly(A)⁺ RNA, and a 203-bp DNA fragment was selectively amplified by polymerase chain reaction (PCR) (see Materials and methods). DNA sequence analysis revealed a single open reading frame (ORF) that happened to encode the 20-amino-acid G542 peptide (Table 1), as well as the G537 and G538 peptides to which the degenerate oligodeoxynucleotides were designed. Thus, the 203-bp DNA fragment specified three of the six TCF-1 α peptides obtained by protein microsequencing, strongly suggesting that the amplified DNA was derived from a portion of the TCF-1 α gene.

The 203-bp PCR fragment was then used to screen a λ ZAP II Jurkat library for larger cDNA clones. A total of 80 positive clones were detected from a screen of 250,000 individual plaques, indicating that TCF-1 α is encoded by a moderately abundant mRNA class [0.03% of total Jurkat poly(A)⁺ RNA]. The largest of 12 cDNA clones isolated by in vivo rescue (clone 3.2) was 3.2 kb in length, in close agreement with the size observed for the largest mRNA (see below). The cDNA insert was sequenced on both DNA strands, using full-length and ExoIII-deleted templates in combination with overlapping oligodeoxynucleotide primers. In addition, a second cDNA insert

Table 1. Amino acid sequence of tryptic TCF-1 α protein fragments

Peptide	Position ^a	Residues ^b	Sequence ^c	Probes ^d
G537	89–104	16	EHPDDGKHPDGGLYNK	5'-GG RTG YTT NCC ATC ATC RGG RTG YTC-3'
G538	37–50	14	IFAEISHPEEGDL	5'-CAY CCY GAR GAR GAR GGN GA-3'
G539	348–353	6	YYELAR	
G540	142–159	18	VPVQPS(H)AVHPLTPLIT	
G541	383–391	9	LQESASGTG	
G542	55–74	20	SSLVNESEIIPASNGHEVAR	

Purified peptides were analyzed by Edman degradation on an Applied Biosystems Protein Sequencer 470A, equipped with an on-line PTH Analyzer 120A. Recovery yields were between 33.6 and 1.5 pmoles of PTH amino acid per cycle.

^aThe location of each peptide within the coding sequence (amino acid is assigned to the putative initiator methionine).

^bThe number of residues sequenced in each peptide.

^cA single histidine residue in peptide G540 (in parenthesis) was not assigned by sequencing.

^dSequence of the degenerate oligonucleotides designed to encode peptides G537 and G538, which were used for PCR amplification.

(clone 3.1) was characterized and found to differ from the original cDNA by a short in-frame deletion, as described below.

Structural features of the TCF-1 α protein predicted by the cDNA sequence

Inspection of the sequence of the largest TCF-1 α clone (clone 3.2) revealed a single extended ORF capable of encoding a 399-amino-acid protein, preceded by an unusually long (655-bp) 5'-untranslated leader segment and followed by a 1.211-kb 3' noncoding segment containing a canonical polyadenylation site (shown schematically in Fig. 1C). An initiator methionine codon that fits well with the Kozak consensus sequence (Kozak 1990) is present starting at position +656 in the cDNA. The ORF continues for 106 amino acid residues upstream of this methionine before a stop codon is reached in the 5'-untranslated leader region (UTR). The predicted ORF contains all six sequenced peptides from the 55-kD TCF-1 α protein (Fig. 1C). Southern analysis confirmed that TCF-1 α is encoded by the human genome and suggests that the protein is encoded by a single gene (data not shown).

Interestingly, the predicted molecular mass of the TCF-1 α protein (44.2 kD) is at least 10 kD smaller than the apparent mass of the native TCF-1 α protein. Therefore, either the TCF-1 α protein migrates aberrantly in denaturing SDS gels or the putative initiator methionine at +656 is not the first amino acid of the synthesized protein. For example, translation of TCF-1 α could initiate at either of two leucine residues (CUG codons) located upstream of the first methionine (AUG codon), as has been observed with certain viral proteins (for review, see Kozak 1990). To distinguish between these two possibilities, we compared the apparent size of proteins synthesized in vitro from constructs that either contain or lack the 5' UTR. In the construct lacking the leader, translation initiation was specified by a 21-amino-acid leader segment of the bacterial gene 10 protein. A single TCF-1 α protein, with an apparent molecular mass of 55 kD, was synthesized from each of these constructs (Fig. 2). Thus, translation of the TCF-1 α protein must begin at the predicted initiator methionine (+656) to yield a 399-amino-acid protein that migrates in SDS-PAGE substantially larger than its predicted size. In addition, RNA derived from the construct lacking the 5' UTR was translated at substantially higher levels than RNA derived from the full-length clone, indicating that the long GC-rich 5' leader region of the TCF-1 α gene is highly inhibitory to translation in vitro. The discrepancy between the predicted and actual size is due to the particular composition of the amino-terminal two-thirds of the TCF-1 α protein, which is exceptionally rich in helix-distorting residues (14% Pro), since fragments expressed from the carboxy-terminal one-third of the protein were found to migrate according to their predicted size (see Fig. 5, below).

The protein sequence predicted from the ORF of the largest TCF-1 α cDNA isolate (clone 3.2) is presented in

Figure 2. In vitro synthesis of TCF-1 α proteins. (Left) SDS-PAGE analysis of 35 S-labeled proteins synthesized in translation reactions that lack added RNA (lane 1) or contain either 2 or 4 μ l (lanes 2 and 3, respectively) of in vitro-synthesized TCF-1 α RNA. Lane M contains 14 C-labeled protein standards (Amersham). (Right) SDS-PAGE analysis of proteins generated from reactions with no added RNA (lane 1), with RNA from the entire 3.2-kb TCF-1 α clone (+ leader, lane 2), or with RNA from a bacterial TCF-1 α fusion protein that lacks the 5'-untranslated leader (- leader, lane 3).

Figure 3. We have also characterized a shorter cDNA isolate, clone 3.1, which differs from the full-length cDNA in that it contains a precise 28-amino-acid in-frame deletion (amino acids 214–241; see Fig. 3A), which removes a short, serine- and threonine-rich segment of the protein. The first notable feature suggested by the sequence of these clones is that the TCF-1 α proteins are particularly rich in proline (12%) and serine (10%) residues. A basic region between amino acids 374 and 382 might serve as a nuclear localization signal. An additional domain from amino acids 244–286 is notably rich in proline and histidine residues (40% proline plus histidine) and may be weakly related to the “paired repeat” domains near the amino termini of the *Drosophila* homeotic *ems* and *bicoid* genes (Burri et al. 1989). Most strikingly, the carboxy-terminal one-third of the protein encompasses a 68-amino-acid domain (amino acids 295–362) that strongly resembles a conserved region of HMG and nonhistone chromosomal proteins (NHPs). This motif is also present in the *Schizosaccharomyces pombe* mating-type compatibility gene, Mc (Kelly et al. 1988), the human RNA polymerase I enhancer-binding protein hUBF (Jantzen et al. 1990), and SRY, the putative mammalian testis-determining factor (Gubbay et al. 1990; Sinclair et al. 1990). Although the HMG protein family is thought to associate relatively nonspecifically with DNA, only slightly preferring AT-rich DNA segments, the hUBF transcription factor does bind specifically to an extended GC-rich segment of the RNA polymerase I enhancer (Jantzen et al. 1990). Deletion of these HMG repeats prevents stable attachment to DNA affinity matrices, suggesting that the HMG motifs mediate specific binding of hUBF to DNA.

At least 33 residues within the 68-amino-acid HMG motif are identical or highly conserved among the previously characterized members of the HMG protein family

61 AAC ATC TAA AAA GGG ACA TCT CCA CGG TGG GTG GCT OCT TTT TCT TTT TCT TTT TTT OCC
 121 AGC CTT CAG GAA GTG GAC TGT TCG TTA TCT TCT OCT TGC ACC TTC TTT TGG GGA AAC
 121 GGG GOC CTT CTG CCC AGA TCC OCT CTC TTT TCT CGG AAA ACA AAC TAC TAA GTC GGC ATC
 181 CGG GGT AAC TAC TGT GCA GAG GGT TTC CGC GGA GAC GCG CCC GCG OCT OCT CTG CAC
 241 TTT GGG GAG GCG TGC TOC CTC CAG AAN CGC CGT TCT OCT GCG GCA AAT CCC GGC GAC GCG
 301 GGG TCG CGS GGT CGC CGC CGG CGC AGC CTC CTC TAG CGC GCG CCC GCG AGA CGC CCC GSG
 361 AGT CGC CAG GTA CCG CAG CCC TCG CCC OCT AGT CGC CTT GCG CCT GGG GCG GCG CGC CTG
 421 OCT CGS TCT CCG ACA AGC GGG AAA CGC GCG GCG CGC GCG GGA TTT GGG GCG CGC GCG AGC
 481 TGC TCG TCG CCG CGC CGC CGC CGC CGC TCG CCC GCG CCT GCG CCT GCG OCT GTC CTG
 541 CTG CAC GAA CCC TTC CAA CTC TCC TTT OCT CCG ACA CCC TTG AGT TAC COC TCT GTC TTT
 601 CCT GCT GTT GCG GCG GTG CTC CAG CGC AGC GGA GAT TAC AGA GCC GCG GAG ATT OCC
 MET pro 2

661 CAA CTC TCC GGA GSA GGT GGC GGC GGC GGG GGG GAC CCG GAA CTC TGC GGC ACG GAC GAG
 gln leu ser gly gly gly gly gly gly gly asp pro glu leu cys ala thr asp glu 22

721 ATG ATC CCC TTC AAG GAC GAG GGC GAT OCT CAG AAG GAA AAG ATC TTC GOC GAG ATC AGT
 MET ile pro phe lys asp glu gly asp pro gln lys glu lys ile phe ala glu ile ser 42

781 CAT CCC GAA GAG GAA GGC GAT TTA OCT GAC ATC AAG TCT TCC TTG GTG AAL GAC TCT GAA
 his pro glu glu gly asp leu ala asp leu lys ser ser val asn glu ser glu 62

841 ATC ATC CCG GGC AGC AAN GGA CAC GAG GTG GCC AGA CAA GCA CAA ACC TCT CAG GAG CCC
 ile ile pro ala ser asn gly his glu val ala arg gln ala gln thr ser gln glu pro 82

901 TAC CAC GAC AAG GCC AGA GAA CAC CCC GAT GAC GGA AAG CAT CCA GAT GGA GGC CTC TAC
 tyr his asp lys ala arg glu his pro asp asp gly lys his pro asp gly gly leu tyr 102

961 AAC AAG GGA CCC TCC TAC TCG AGT TAT TCG GGG TAC ATA ATG ATG CCA AAT ATG AAT AAC
 asn lys gly pro ser tyr ser ser tyr ser gly tyr ile met met pro asn met asn asn 122

1021 GAC CCA TAC ATG TCA AAT GGA TCT CTT TCT CCA CCC ATC CCG AGA ACA TCA AAT AAA GTG
 asp pro tyr met ser asn gly ser leu ser pro pro ile pro arg thr ser asn lys val 142

1081 CCC GTG GTG CAG CCA TCC CAT GCG GTC CAT OCT CTC ACC CCC CTC ATC ACT TAC AGT GAC
 pro val val gln pro ser his ala val his pro leu thr pro leu ile thr tyr ser asp 162

1141 GAG CAC TTT TCT CCA GGA TCA CAC CCG TCA CAC ATC CCA TCA GAT GTC AAC TCC AAA CAA
 glu his phe pro pro ser his pro ser his ile pro ser asp val asn lys gln 182

1201 GGC ATG TCC AGA CAT OCT CCA GCT OCT GAT ATC OCT ACT TTT TAT CCC TTG TCT CCG GGT
 gly met ser arg his pro pro ala pro asp ile pro thr phe tyr pro leu ser pro gly 202

1261 GGT GTT GGA CAG ATC ACC CCA CCT CTT GGC TGG [CAA GGT CAG OCT GTA TAT CCC ATC ACG
 gly val gly gln ile thr pro pro leu gly trp] gln gly gln pro val tyr pro ile thr 222

1321 GGT GGA TTC AGG CAA CCC TAC CCA TCG TCA GTC GAC ACT TCC ATG TCC AGG] TTT
 gly gly phe arg gln pro tyr pro ser ser ser val asp thr ser met ser arg] phe 242

1381 TCC CAT CAT ATG ATT CCC GGT OCT OCT GGT CCC CAC ACA ACT GGC ATC OCT CAT CCA GCT
 ser his his met ile pro gly pro pro gly pro his thr thr gly ile pro his pro ala 262

1441 ATT GTA ACA CCT CAG GTC AAA CAG GAA CAT CCC CAC ACT GAC AGT GAC CTA ATG CAC GTG
 ile val thr pro gln val lys gln gly his pro his thr asp ser asp leu met his val 282

1501 AAG CCG CAG CAT GAA CAG AAG AAG GAG CAG GAG CCA AAA AGA CCT CAC ATT AAG AAG OCT
 lys pro gln his glu gln arg lys glu gln glu pro lys arg pro his ile lys lys pro 302

1561 CTG AAT GCT TTT ATG TTA TAC ATG AAA GAA ATG AGA GCG AAT GTC GTT GCT GAG TGT ACT
 leu asn ala phe met leu tyr met lys glu met arg ala asn val val ala glu cys thr 322

1621 CTA AAA GAA AGT GCA GCT ATC AAC CAG ATT CTT GGC AGA AGG TGG CAT GOC CTC TCC OCT
 leu lys glu ser ala ala ile asn gln ile leu gly arg arg trp his ala leu ser arg 342

1681 GAA GAG CAG GCT AAA TAT TAT GAA TTA CCA CGG AAA GAA AGA CAG CTA CAT ATG CAG CTT
 glu glu gln ala lys tyr tyr glu leu ala arg lys glu arg gln leu his met gln leu 362

1741 TAT CCA GGC TGG TCT GCA AGA GAC AAT TAT GGT [AAG AAA AAG AAG AGG AAG AGA GAG AAA
 tyr pro gly trp ser ala arg asp asn tyr gly] lys lys lys lys arg lys arg glu lys 382

660 GENES & DEVELOPMENT

(Fig. 4). Overall, TCF-1 α contains 24 of the 33 shared residues and conserves 8 additional amino acid residues with the Mc and SRY proteins that are not found in the hUBF, NHP-6, and HMG-1 proteins. Only two residues are not found in TCF-1 α that are otherwise invariant among other members of the family: a proline residue at position 26 (threonine in TCF-1 α) and a basic amino acid at position 49 (glutamine in TCF-1 α). Thus, TCF-1 α is most similar overall to the Mc and SRY proteins. The Mc, SRY, and TCF-1 α proteins are distinguished further from hUBF by the fact that each contains only a single HMG box, whereas this motif is repeated at least three times in hUBF.

Chou-Fasman analysis of the TCF-1 α protein sequence predicts that the HMG domain forms two extended 26-amino-acid α -helical stretches separated by a 6-amino-acid spacer segment (Fig. 4). The hUBF protein has also been noted to contain a high α -helical propensity within its repeated HMG motifs (Jantzen et al. 1989). α -Helical repeats are found in a variety of DNA-binding domains, including those of the helix-loop-helix, helix-turn-helix, and homeo box proteins (for review, see Johnson and McKnight 1989; Mitchell and Tjian 1989). Interestingly, the second half of the HMG domain displays weak homology to a portion of the DNA-binding domain of the *c-ets* oncogene (Fig. 4), suggesting that the HMG and *ets* DNA-binding protein families could be very distally related.

A single HMG motif mediates specific binding of TCF-1 α to DNA

To compare the DNA-binding properties of the cloned and native TCF-1 α proteins, we subcloned the TCF-1 α gene (clone 3.2) into a modified pGEMEX vector that utilizes the phage T7 RNA polymerase promoter (Fig. 5A; H. Mangalam, pers. comm.). TCF-1 α protein was then expressed in a bacterial strain (pLys-s) in which the

T7 RNA polymerase is induced upon treatment with IPTG (Studier et al. 1990), and soluble bacterial fractions enriched for the induced TCF-1 α protein were prepared (Fig. 5B). The induced bacterial TCF-1 α protein (bTCF) bound the human TCR α enhancer in a manner identical to that of the purified native TCF-1 α protein (hTCF) in DNase I footprint experiments (Fig. 5C). No specific binding to this DNA fragment was observed with uninduced bacterial extracts (data not shown). A double point mutation in the TCR α enhancer destroyed the binding of bTCF in mobility-shift experiments (Fig. 5D). We have shown previously that this mutation eliminates the binding of the purified TCF-1 α protein in vitro and dramatically reduces TCR α enhancer activity in vivo (Waterman and Jones 1990). Thus, the bacterially expressed cloned protein maintains the binding specificity observed with the native protein purified from Jurkat cells.

To map the DNA-binding domain of the TCF-1 α protein, two mutant cDNAs were constructed and expressed using these bacterial vectors. One construct, Δ HMG, contains an in-frame deletion from amino acids 237 to 396 and, hence, encodes a protein lacking the HMG domain. The second construct, called HMG, generates a truncated 162-amino-acid protein (amino acids 237–399) derived from the carboxy-terminal third of the protein that contains the HMG motif. Soluble extracts containing induced bacterial TCF-1 α proteins were prepared and analyzed for enhancer-binding activity. The DNA-binding properties of the truncated HMG-containing protein were found to be indistinguishable from those of the full-length TCF-1 α protein in DNase I footprint (Fig. 5C) and Southwestern blot (Fig. 5F) experiments. In contrast, the Δ HMG protein did not bind the TCR α enhancer in either footprint, mobility-shift, or Southwestern blot experiments (Fig. 5F, and other data not shown). The Southwestern blotting procedure detects only specific protein-DNA interactions, because the renatured TCF-1 α proteins were not bound by a la-

Figure 4. Comparison of different HMG domain sequences. Shown is a 68-amino-acid region conserved among the following proteins: (HU TCF-1 α) Human TCF-1 α ; (YST Mc) mating-type protein Mc from *S. pombe*; (MU-SRY) murine SRY; (HU-SRY) human SRY; (YST NHP-6) yeast nonhistone protein 6; (BOV-HMG-1) calf thymus histone mobility group protein-1; (HU-UBF-1, -2, -3) three HMG box repeats of hUBF. Highly conserved amino acids are shaded. TCF-1 α residues that are conserved only with the Mc and SRY proteins are shown in open boxes. A summary of the amino acid conservation is shown at top, where amino acids found in at least five of the nine HMG proteins are boxed; residues shared only among TCF-1 α , Mc, and SRY are circled. Dashes indicate nonconserved residues; (B) basic amino acids; (ψ) hydrophobic amino acids. Two potential α -helices in the HMG region of TCF-1 α are indicated above the sequence. A region of the DNA-binding domain of the murine *c-ets*-1 proto-oncogene that shows limited homology to the HMG motif is shown at bottom. Asterisks (*) denote tryptophan repeats that are conserved among the *c-ets* protein family.

Figure 5. DNA-binding properties of TCF-1 α proteins expressed in bacteria. (A) Schematic illustration of bacterial TCF-1 α expression vectors. (FL) Full-length protein expressed from the 3.2-kb cDNA; (Δ HMG) mutant protein lacking amino acids 237–396; (HMG) the truncated HMG-containing protein (amino acids 237–399). (B) SDS-PAGE analysis of the induced bacterial FL (55-kD) and HMG (21-kD) TCF-1 α proteins. (M) Protein markers. Aliquots containing 10 μ l of total lysed bacterial extract obtained before (–) or after (+) induction with 0.5 mM IPTG (Materials and methods) were analyzed by 10% SDS-PAGE and visualized by silver staining. (C) DNase I footprint analysis of affinity-purified Jurkat TCF-1 α cells (native; 0.6 and 1.5 ng) or bacterial extracts containing the FL TCF-1 α (1.7 and 3.4 μ g) or truncated HMG-containing TCF-1 α (0.5 and 1.0 μ g) proteins, respectively. (D) Gel mobility-shift analysis of the binding of bacterial TCF-1 α (0.25 μ g; truncated HMG protein) to the wild-type TCR α enhancer (WT) or to a double point mutant in the enhancer (see Materials and methods). (E) (●) SDS-PAGE analysis of the induced bacterial FL (55-kD), Δ HMG (27-kD), and HMG (21-kD) TCF-1 α proteins. Note that both the FL and Δ HMG proteins migrate aberrantly on SDS-PAGE, whereas the shorter HMG protein migrates according to its predicted size. (F) Southwestern analysis of bacterial TCF-1 α proteins. Proteins were transferred from 10% SDS-PAGE to nitrocellulose and probed with oligonucleotides of the wild-type TCR α enhancer (left, wild type) or the double point mutant (right, double point mutation). (M) 14 C-labeled molecular weight markers (Amersham).

beled DNA probe containing a double point mutation in the TCR α enhancer (Fig. 5F).

At present, little is known about how sequence-specific HMG proteins interact with their DNA targets. To investigate whether the HMG proteins bind to DNA as dimers or other multimers, the truncated and full-length bacterial TCF-1 α proteins were mixed and the resultant protein–DNA complexes were analyzed by mobility-shift experiments. Complexes specific for the truncated and full-length proteins formed; however, we failed to observe any complexes of intermediate size that would

suggest the presence of mixed heterodimers, even when the mixture containing the two forms of the protein was denatured with guanidine-HCl (to dissociate any subunits) and then renatured prior to the binding assay (data not shown). We were unable to assess the binding of short and long proteins that had been cotranslated because TCF-1 α proteins formed by *in vitro* transcription and translation are unable to bind specifically to DNA (data not shown). We conclude that TCF-1 α most likely binds to DNA as a monomer, although it remains possible that it forms a tightly associated dimer or other

multimer that cannot be dissociated or reassembled following translation.

Thymus- and T cell-restricted expression of TCF-1 α mRNA

We demonstrated previously that TCF-1 α is a T cell-specific DNA-binding protein, present in nuclear extracts from Jurkat and CCRF-CEM T cell lines but not in extracts of a mature B cell line (JY) or the nonlymphoid HeLa cell line (Waterman and Jones 1990). To determine whether this cell-type specificity results from transcriptional regulation of the TCF-1 α gene, we analyzed various cell lines and tissues for TCF-1 α mRNA by using Northern blot experiments (Fig. 6). A DNA probe consisting of the original PCR primer detected two TCF-1 α mRNAs: an abundant species of 3.4 kb and a minor mRNA of 2.3 kb (Fig. 6A). TCF-1 α mRNA was detected in total RNA prepared from Jurkat cells (a mature helper T cell), CCRF-CEM cells (an immature T cell line), and a subclone of CCRF-CEM (clone 1—CEM: an activated CEM line) but was not present in JY (an Epstein-Barr virus (EBV)-transformed B cell line) or U937 cells (a macrophage cell line). TCF-1 α -specific mRNA was also present in the murine WEHI T-cell line and was found to be absent in the following murine B cell lines: S107, P3, MOPC, CH12, and J558 cells (data not shown). A very low level of TCF-1 α mRNA could be detected in the

Namalwa B cell line (data not shown); however, TCF-1 α DNA-binding activity was not observed in nuclear extracts from Namalwa cells. Thus, it appears that the low level of RNA synthesized in the Namalwa B cell line is insufficient to generate substantial levels of TCF-1 α protein or, alternatively, the gene may be regulated post-transcriptionally in these cells. Overall, the TCF-1 α gene appears to be regulated transcriptionally in a highly T cell-specific manner.

In addition, we noted that the immature CCRF-CEM T cell line expresses at least fivefold lower levels of TCF-1 α mRNA than are found in the mature Jurkat cell line (Fig. 6A, cf. lanes 1 and 2). A subclone of this CEM cell line (clone 1), which differs from the parental cell line in that it possesses many of the features of an activated cell, including the expression of endogenous TCR α on the cell surface (Shackelford et al. 1987), was found to contain TCF-1 α mRNA at levels comparable to those found in Jurkat cells (Fig. 6A, cf. lanes 2 and 3). Similarly, TPA treatment of the uncloned parental CEM cell line increased TCF-1 α mRNA levels three- to fourfold (data not shown). We conclude that TCF-1 α mRNA is present in both mature (Jurkat, WEHI) and immature (CEM) T cell lines and can be induced upon activation of the immature CEM cell line.

To better define the overall pattern of expression of TCF-1 α mRNA, a DNA fragment corresponding to the HMG domain was used to probe a Northern blot containing RNA from different murine tissues, derived primarily from 4-week-old mice (Fig. 6B). Two murine TCF-1 α mRNAs were detected that are similar in size to the human TCF-1 α mRNAs, although the 2.3-kb mRNA species was relatively more abundant in murine cells. The two TCF-1 α mRNAs were present at high levels in the thymus and were not observed in RNA derived from the brain, heart, lung, kidney, liver, or spleen. Thus, TCF-1 α mRNA is expressed in a highly specific manner that is restricted to the thymus. The absence of significant TCF-1 α mRNA in the spleen is consistent with the relative lack of expression of this gene in mature B cell lines.

Figure 6. TCF-1 α mRNA expression is restricted to T cells. (A) Aliquots of 15 μ g of total RNA purified from the different indicated cell lines were separated on a 1.0% formaldehyde-agarose gel, transferred to Hybond N membrane, and probed with a random-primed DNA fragment corresponding to the original PCR probe (amino acids 15–82; see Fig. 1C). Arrows indicate the sizes of the two TCF-1 α -specific mRNAs and their migration position relative to 18S and 28S RNAs. Size estimates were determined relative to a series of RNA marker standards that were run in parallel. (B) Aliquots of 20 μ g of total RNA from different murine tissues were separated on a 1% formaldehyde-agarose gel, transferred to Hybond N, and probed with a random-primed 450-bp DNA probe encompassing the HMG domain (amino acids 253–399). Ethidium bromide staining of the 18S rRNA band is shown below each panel to indicate total RNA levels in each lane.

Transcriptional activation by the cloned TCF-1 α protein

The TCF-1 α protein encoded by the full-length cDNA (clone 3.2) was tested for transcriptional activity in the HeLa nonlymphoid cell line, which we have shown previously to lack endogenous TCF-1 α DNA-binding activity. For these experiments, a construct containing a single copy of the 95-bp core domain of the human TCR α enhancer positioned upstream of the HSV-1 TK promoter and luciferase gene (α 1tk-LUC; Waterman and Jones 1990) was used as the reporter gene. Expression of the TCF-1 α protein was obtained by inserting the 3.2 cDNA clone (the entire ORF and 161 bp of leader sequences) at position –19 of the human cytomegalovirus (hCMV) promoter (–1122 to –19). As a control, the α 1tk-LUC reporter was cotransfected with pCMV-1 (the expression vector lacking the TCF-1 α gene) and the data

are presented as fold activation seen with pCMV–TCF-1 α , relative to pCMV-1. We observed a strong (20-fold) activation of the α 1tk–LUC vector upon cotransfection with the pCMV–TCF-1 α expression plasmid in HeLa cells (Fig. 7). This activation was slightly lower (sixfold) at higher ratios of the pCMV–TCF-1 α expression vector to the reporter gene, most likely due to competition from the strong hCMV promoter. No increase in luciferase activity was observed upon cotransfection with pCMV-1. Furthermore, the cloned TCF-1 α protein was unable to activate constructs in which the wild-type TCR α enhancer was substituted with an enhancer containing a double point mutation in the TCF-1 α -binding site (α m1tk–LUC; Fig. 7), which destroys the binding of TCF-1 α (Fig. 5C). Therefore, *trans*-activation in this non-lymphoid cell line requires the TCF-1 α protein as well as an intact TCF-1 α -binding site.

Discussion

Among the key regulatory proteins important for T lymphocyte formation are likely to be transcription factors that activate and restrict the expression of cell-surface antigens associated with specialized T cell subtypes. TCF-1 α , a T cell-specific transcription factor that activates the human T cell receptor C α enhancer, is a good candidate for such a factor (Waterman and Jones 1990). Indeed, TCF-1 α is likely to play a varied role in T cell-specific gene expression because it binds to transcription control regions of antigen genes that are transcribed at both early (e.g., p56^{lck}, CD3, and TCR δ), and relatively late stages of T lymphocyte development (TCR α , CD4; M.L. Waterman, K. Jones, J. Siu, and S. Hedrick, unpubl.). To further characterize this transcription factor, we isolated the gene for TCF-1 α , using amino acid sequence information derived from the purified Jurkat TCF-1 α protein. Analysis of two related TCF-1 α cDNAs revealed that the amino-terminal two-thirds of the predicted

TCF-1 α protein is rich in proline, serine, and threonine residues, whereas the carboxyl terminus is strikingly devoid of proline residues and contains a 68-amino-acid domain conserved among the HMG protein family. The two TCF-1 α cDNAs differ from one another by a short in-frame deletion that removes a serine- and threonine-rich segment that lies upstream of the HMG domain. Comparison of these and other partially characterized cDNAs suggests that at least some of the diversity of the 57- to 53-kD TCF-1 α protein family may be generated by alternative splicing. Northern blot analysis reveals that TCF-1 α is encoded by two discrete mRNA species that are highly restricted to T cells, the larger of which correlates well with the size of the two characterized TCF-1 α cDNAs. It will be interesting to learn the structure of cDNAs representative of the shorter transcripts and whether the variant cDNA forms encode proteins with altered activity. The strong tissue specificity of expression, relative abundance of possible phosphorylation sites, and potential for splicing indicate that TCF-1 α activity could be regulated both transcriptionally and post-transcriptionally in T lymphocytes.

One of the more striking findings presented here is that TCF-1 α is a member of the HMG protein family. Until recently, it was thought that this family consisted solely of abundant chromosomal proteins that associate nonspecifically with DNA; however, the discovery that the hUBF RNA polymerase I transcription factor contains several HMG motifs (Jantzen et al. 1990) clearly demonstrated that at least some HMG proteins are involved in more complex regulatory phenomena. More recently, the cDNA for *SRY*, a candidate for the human testis determining factor, was also found to contain an HMG domain (Gubbay et al. 1990; Sinclair et al. 1990). Although specific gene targets for *SRY* have not yet been identified, the regulatory nature of the gene product strongly suggests that it is a specific DNA-binding protein. The reiterated HMG motifs of hUBF were required

Figure 7. Transcriptional activation by the cloned TCF-1 α protein in nonlymphoid cells. Reporter constructs for transfection analysis contained a single copy of the 95-bp core TCR α enhancer (α 1tk) upstream of the TATA box of the HSV-1 thymidine kinase (tk) promoter and bacterial luciferase gene (Luc), or a single copy of the enhancer bearing the double point mutation that eliminates the binding of TCF-1 α (α m1tk; Waterman and Jones 1990), as indicated schematically at top. These constructs were cotransfected into HeLa SL-3 cells along with an expression vector in which TCF-1 α protein (the entire 3.2-kb clone) was positioned downstream of the cytomegalovirus (CMV) immediate early gene promoter. The pCMV-1 promoter plasmid lacking an insert was used as a control plasmid in separate cotransfection experiments. Luciferase activity from the reporter gene was measured in light units, normalized by adjustment for protein, and the data are expressed as the fold induction observed relative to the luciferase reporter plasmid cotransfected with a pCMV-1 control plasmid. Shown are representative data obtained with two different ratios of TCR α –luciferase reporter and TCF-1 α expression plasmids, as indicated at the left of the graph.

for its attachment to DNA affinity resins, implicating the HMG motif in DNA recognition (Jantzen et al. 1990). The DNA-binding properties of TCF-1 α differ significantly from those of hUBF in that the former recognizes a short conserved pyrimidine-rich element (Table 2), whereas hUBF binds relatively weakly to an extensive GC-rich region of DNA, and different characterized hUBF-binding sites lack an obvious consensus motif (Learned et al. 1986). Indeed, the unusual binding properties of hUBF have led to the proposal that specific binding may require elements of DNA structure as well as sequence (Jantzen et al. 1990). A second notable difference is that the primary structure of hUBF contains three or four repeated HMG boxes, whereas TCF-1 α , like the Mc and SRY gene products, contains only a single HMG motif. Expression studies reveal that the single HMG box of the TCF-1 α gene is both necessary and sufficient for specific binding to its pyrimidine-rich target site in DNase I footprint, mobility-shift, and Southwestern experiments (Fig. 5), and mixing experiments indicate that the protein may bind as a monomer. Overall, the different members of the HMG protein family display a remarkable diversity in protein-DNA interactions, ranging from nearly nonspecific to highly specific binding, which appears to be determined by the exact composition of the HMG box.

The HMG domain of TCF-1 α is predicted by Chou-Fasman analysis to consist of two large α -helical segments separated by a short spacer (Fig. 4) which, overall, is somewhat reminiscent of the helix-turn-helix and helix-loop-helix DNA-binding structures. The α -helices in the TCF-1 α -binding domain are considerably longer, however, and are not notably amphipathic in nature. Shorter α -helical stretches are also present in hUBF (Jantzen et al. 1990) and may be an important structural component of the HMG-binding domain. The differences in the DNA-binding properties of TCF-1 α and hUBF noted above may be specified by the distinct number of HMG boxes present in each protein, or, more likely, by individual amino acid differences within the HMG domains. In this regard, it is interesting to note that TCF-1 α shares an additional 8 residues with the yeast Mc and mammalian SRY proteins that are not found in hUBF or the other members of the HMG protein family. This region of extended similarity is localized to both extremes of the HMG domain and may define a subclass of the HMG family that is related to DNA-binding sequence preference or to the degree of DNA-binding

specificity. A recent report that some XY females contain point mutations in the amino-terminal half of the SRY HMG box (Berta et al. 1990) indirectly underscores the importance of this region of the HMG motif. Further identification of important amino acid residues in the HMG domain awaits site-directed mutagenesis of the HMG box, as well as identification of specific gene targets for the Mc and SRY proteins. At present, we conclude from the analysis of TCF-1 α that a single HMG motif can be sufficient to interact with a specific target and that members of the HMG family are critical not only for transcription of RNA polymerase I genes and sex determination events but also for cell- and tissue-restricted activation of RNA polymerase II enhancers.

Interestingly, our computer search also indicated that the HMG motif is distally related to the DNA-binding domain of the *c-ets* proto-oncogene family (for review, see Karim et al. 1990). In contrast to the members of the HMG family, however, *c-ets* contains only 3 of 7 invariant residues, and the homology is limited to the distal half of the HMG box (Fig. 4). Contiguous with this region of *c-ets* lies a series of three "tryptophan repeats" that resemble the binding motif of *c-myb* and have been proposed to be an important component of the ETS DNA-binding domain (Karim et al. 1990). These repeated tryptophan residues are not conserved in the HMG box, and there is no striking similarity between the *c-ets*- and TCF-1 α -binding sites, apart from the fact that both are short, pyrimidine-rich elements. Nonetheless, the *c-ets* proteins are also important transcription factors in lymphoid cells (Bosselut et al. 1990; Klemsz et al. 1990; Waslyk et al. 1990), and it will be interesting to learn whether the modest homology described here is reflected in the manner in which these two different classes of proteins interact with DNA.

Unlike the elements for other tissue-specific activators, reiterated copies of the TCF-1 α -binding site do not create an enhancer, or minimal unit capable of independent enhancer activity in T cells (Waterman and Jones 1990). Instead, it appears that the strong activity of TCF-1 α in T lymphocytes requires interactions with neighboring DNA-binding proteins. For example, the minimal TCR α enhancer appears to absolutely require contiguous binding sites for three proteins: CREB (or a related CRE-binding protein), TCF-1 α , and a second T cell-specific factor, TCF-2 α . Ho et al. (1990) have recently shown that the *c-ets-1* protein can bind specifically to the TCF-2 α region of the enhancer and that the *ets*-binding sequence is critical for enhancer activity in vivo. Although these data suggest that TCF-2 α is identical to *c-ets-1*, several observations indicate that recognition of the TCF-2 α site is more complex. First, TCF-2 α -binding activity appears to be restricted to T cells, whereas *c-ets-1* is abundant in B cells and macrophages, as well as T cells (Chen 1985; Bhat et al. 1989). In addition, Southwestern experiments indicate that TCF-2 α is a 63-kD protein (M.L. Waterman, unpubl.), substantially larger than the 54-kD *c-ets-1* protein. One possibility is that a T cell-specific protein binds in conjunction with *c-ets-1* to the TCF-2 α -binding site or, al-

Table 2. Sequence comparison of TCF-1 α DNA-binding sites

TCR α	5'-GGCAC <u>CCCTTT</u> GAA-3'
TCR δ	5'-AAAGC <u>CCCTTT</u> GAA-3'
HIV-1	5'-ACCAGTCTTTGTA-3'
<i>lck</i>	5'-GGCCT <u>CTCTG</u> TGAA-3'

Shown are sites from various genes that are recognized with relatively high affinity by affinity-purified TCF-1 α protein (Waterman and Jones 1990). The pyrimidine-rich core motif conserved among these sites is underlined.

ternatively, TCF-2 α might be a distinct, T cell-specific member of the *c-ets* protein family.

In contrast to *c-ets*, the expression of TCF-1 α mRNA is highly restricted to the thymus and T cell lines and was not found in B cell lines, macrophages, or the spleen. Somewhat lower levels of TCF-1 α mRNA were detected in immature CCRF-CEM cells than in mature T cell lines, consistent with the low but detectable activity of the TCR α enhancer in these cells, and these low mRNA levels were increased in an activated CEM subclone as well as in TPA-treated CEM cells. Although transcriptional induction of the TCF-1 α gene could contribute to the increased activity of the α -chain enhancer in TPA-treated CEM cells, it is unlikely to be the sole determinant of TCR α induction because enhancer mutants that can no longer bind TCF-1 α retain the ability to be induced by TPA (Waterman and Jones 1990). Indeed, it is likely that the TCF-2 α /*ets* proteins are also induced upon activation of immature T cells (Ho et al. 1990) and could function in combination with TCF-1 α to activate the enhancer. The finding that TCF-1 α mRNA is present in immature T cells and in the thymus of young mice is consistent with its potential role in regulating the expression of genes expressed at relatively early times in T cell development.

Given the strong context dependence of transcriptional activation by TCF-1 α in T cell lines, it was somewhat surprising to find that the cloned TCF-1 α protein is capable of activating minimal promoters bearing a single copy of the TCR α enhancer in nonlymphoid (HeLa) cells (Fig. 7). This strong activation of the TCR α enhancer was eliminated by a double point mutation that destroys the TCF-1 α -binding site. These results indicate that TCF-1 α can act alone in these cells to induce transcription in a binding site-dependent manner. In contrast, the activity of the TCR α enhancer in T cell lines was completely dependent on each of the three protein-binding sites (CREB, TCF-1 α , and TCF-2 α), and partial activity was not observed with constructs bearing only the CRE and TCF-1 α -binding sites. Thus, the requirement for proper protein-protein interactions among the CREB, TCF-1 α , and TCF-2 α proteins appears to be more stringent in T lymphocytes versus nonlymphoid cells. It will be interesting to learn whether TCF-1 α can function synergistically with TCF-2 α /*ets* to induce the TCR α enhancer in nonlymphoid cells.

The differences observed between newly synthesized and endogenous TCF-1 α suggest that the activity of the protein may be modified in lymphocytes to become more dependent on protein-protein interactions. A very similar observation was made for the erythroid-specific GATA-1 regulatory protein, which can activate minimal promoters bearing an upstream GATA-binding site in nonerythroid cells but cannot activate these same promoters in erythroid cell lines (for review, see Orkin 1990). Thus, the endogenous erythroid GATA-1 displays a more restricted requirement for activity than does GATA-1 protein that is newly synthesized in nonerythroid cells, which parallels the differences observed between endogenous and newly expressed TCF-1 α . TCF-1 α

further resembles GATA-1 in that both proteins activate transcription from distal downstream enhancers as well as proximal promoters, are strikingly cell type-specific, and are present at both early and late stages of development. The GATA-1 protein frequently functions in combination with a ubiquitous protein and a cell type-specific AP-1 protein, and its activity is strongly context dependent in erythroid cells (Orkin 1990). These conserved features of the GATA-1 and TCF-1 α proteins may be a hallmark of proteins that function in the early determination of cell type specificity.

A distinct member of the GATA family (GATA-3) has recently been described that is localized to the thymus and brain (Yamamoto et al. 1990) and therefore may also play a role in the transcriptional regulation of T cell-specific enhancers. A potential GATA-binding site is located just outside of the minimal TCR α enhancer core, and although it is not required for the activity of the minimal enhancer in mature T cells, it may play a role in developmental regulation. The converse situation exists in the human TCR δ enhancer, which contains a GATA sequence within the minimal region required for activity in $\gamma\delta^+$ T cell lines (Redondo et al. 1990), and a TCF-1 α -binding site that lies just beyond this region. From the juxtaposition of the TCF-1 α and GATA-3 motifs, it is tempting to speculate that these two proteins play reciprocal roles in the coordinate regulation of the TCR α and TCR δ enhancers in primary thymocytes (Martinez-Valdez et al. 1988) or during T cell development.

The pronounced heterogeneity of the TCF-1 α protein family indicates that different members may play distinct roles in T cell development. Recently, Clevers et al. reported the characterization of a gene, TCF, that was isolated by screening a λ gt11 expression library with the TCF-1 α -binding site from the CD3 ϵ gene enhancer (van de Wetering et al. 1991). Interestingly, the gene is distinct from TCF-1 α , although it demonstrates extensive homology throughout the HMG motif and is expressed predominantly in the thymus. The predicted TCF protein is considerably smaller than TCF-1 α . DNA affinity resins used for the purification of TCF-1 α did enrich for a collection of shorter proteins in the size range of TCF (38–42 kD); however, these proteins bound less avidly to the column and were eluted at lower ionic strength (Waterman and Jones 1990). It will be interesting to compare the DNA-binding specificities, transcriptional activation properties, and developmental pattern of expression of these two proteins directly and to learn whether other genes related to TCF-1 α are also expressed specifically in the thymus.

Together, the CREB, TCF-1 α , and TCF-2 α /*ets* proteins create a powerful enhancer that is critical for tissue-specific expression of the α -chain of the T cell receptor. The striking ability of the TCR α enhancer to regulate transcription over a very large chromosomal domain, which is critical for its normal function in activating distal upstream V α promoters, is also likely to be a contributing factor in the deregulated expression of errantly translocated proto-oncogenes during leukemogenesis (Erikson et al. 1986). Further characterization of these factors, in-

cluding the TCF-1 α HMG protein, will help to determine how a small number of critical proteins interface to create this long-range, cell type-specific enhancer.

Materials and methods

TCF-1 α purification and sequencing

Large-scale DNA affinity purification of TCF-1 α from Jurkat nuclei (1.76×10^{10} cells) was performed as described previously (Waterman and Jones 1990) to yield 16 μ g of TCF-1 α protein at ~50% purity. To separate the TCF-1 α proteins from a single contaminating 116-kD protein, the affinity-purified protein was precipitated with trichloroacetic acid [25% (wt/vol)] and loaded onto a 1-mm 10% SDS-PAGE preparative gel. Following electrophoresis, the proteins were transferred to an Immobilon filter (Millipore) in transfer buffer (TB; 40 mM glycine, 50 mM Tris-HCl, 0.04% SDS, 20% methanol) using an LKB Electrophor electrotransfer apparatus (1 mA/cm² for 1.5 hr at room temperature). The Immobilon membrane was stained with Amido Black (0.1% in methanol/acetic acid/H₂O; 45 : 10 : 45) for 2 min at room temperature, followed by three changes of a destain solution [methanol/acetic acid/H₂O; 450 : 10 : 40 (vol/vol/vol)] and two separate washes with double-distilled water.

Library construction and screening

A λ gt11 expression library was constructed with 5 μ g of poly(A)⁺ RNA isolated from Jurkat cells (Chomczynski and Sacchi 1987). cDNA was made from the poly(A)-selected RNA using a λ ZAP II kit (Stratagene), ligated into λ ZAP II arms, and packaged using Gigapack Gold (Stratagene). The library was found to contain 1.1×10^6 independent recombinant clones and was then amplified before use.

A total of 250,000 individual plaques were transferred to Hybond-N filters (Amersham) and screened with a 203-bp PCR fragment that had been labeled by random priming to a sp. act. of 1.7×10^9 dpm/ μ g. Hybridizations were carried out in 50% formamide, 1% SDS, 5 $\times$ Denhardt's solution, 5 $\times$ SSPE, 20 μ g/ml of salmon sperm, and 1×10^6 cpm/ml of probe, at 42°C for 12 hr. The final wash of filters was in 1 $\times$ SSPE/0.1% SDS, at 65°C for 30 min. Twelve positive plaques were purified through four rounds of plating and rescued into XL1 Blue bacteria by coinfection with the R408 helper phage (Stratagene). In vivo excision of the inserts from the λ ZAP II vector to a Bluescript SK environment was performed according to the manufacturer's suggestions (Stratagene).

PCR amplification

A total of 1 μ g of Jurkat poly(A)⁺ RNA (Chomczynski and Sacchi 1987) was treated with AMV reverse transcriptase (Life Sciences) by priming with 0.5 μ g of oligo(dT) (Promega), according to the protocol of Frohman et al. (1988), except that the reaction was terminated by diluting it fivefold with water, and freezing to -20°C. An aliquot of this cDNA pool (100 ng) was used for each PCR reaction at a final buffer concentration of 10 mM Tris-HCl (pH 8.4), 50 mM KCl, 0.01% gelatin, 3 mM MgCl₂, 200 μ M each dNTP, 1 μ M each degenerate primer, and 2.5 units of Amplitaq polymerase (Cetus). Degenerate oligodeoxynucleotide primers (Table 1) were purified from 15% native polyacrylamide gels prior to use. Each reaction was incubated at 94°C for 3 min and cycled 35 times at 94°C for 1 min, 55°C for 2 min, and 72°C for 3 min, followed by further incubation of the reactions for 5 min at 72°C. Resultant PCR products were analyzed on agarose gels, and the 203-bp fragment was purified from a 1.5% agarose

gel and subcloned into the *Sma*I site of Bluescript SK-. This 203-bp fragment was then sequenced using the T3 and KS (Stratagene) primers.

DNA sequence analysis

The largest insert (3.2 kb) was sequenced with Sequenase enzyme (U.S. Biochemicals) by using a combination of specific oligodeoxynucleotide probes and various ExoIII 3'-deleted templates. Both DNA strands were sequenced within the 5'-UTR and ORF regions of the gene.

Bacterial expression

TCF-1 α cDNA clones and subclones were inserted into a derivative of the bacterial expression vector, pGEMEX (Promega), engineered to remove most of the bacterial gene 10 gene-coding sequences (H. Mangalam, unpubl.). Final clones contained an in-frame fusion with 21 amino acids of gene 10 at the *Hinc*II site in the vector polylinker. All plasmids were sequenced to confirm that the construct and translation frame were correct.

For expression, plasmids were transformed into pLysS bacteria (Studier et al. 1990). Cultures (200 ml) were inoculated with 0.5 ml of an overnight culture and grown at 37°C until bacteria reached a density of OD₆₀₀ = 0.7. TCF-1 α expression was then induced with 0.5 mM IPTG for 2 hr at 37°C. Bacteria were harvested and lysed in 1 $\times$ lysis buffer [LB: 60 mM KCl, 20 mM HEPES (pH 7.9), 1 mM EDTA, 5 mM MgCl₂, 2 mM dithiothreitol (DTT), 0.1 mM phenylmethylsulfonyl fluoride (PMSF), 2 mg/ml of benzamidin, 1 μ g/ml of pepstatin A, 4 μ g/ml of leupeptin, 10 μ g/ml of aprotinin, and 20 μ g/ml of soybean trypsin inhibitor], with two consecutive cycles of freezing and thawing. After a 45-min spin at 50,000 rpm, the supernatant was brought up to 20% glycerol and stored at -100°C. The pellet was resuspended in 4 M guanidine-HCl and placed on a rotating wheel for 30 min at 4°C. Insoluble material was separated by centrifugation for 30 min at 50,000 rpm. The guanidine-HCl was removed from the extract by stepwise dialysis at 4°C. Glycerol was added to 20%, and the extract was centrifuged to remove insoluble protein before freezing.

DNA-binding procedures

Bacterial extracts (amounts indicated in figure legends) were incubated with 1 μ g of salmon sperm DNA in a 14- μ l reaction containing 1 $\times$ EMSA buffer (10 mM HEPES-KOH at pH 7.8, 2.5 mM EDTA, 5 mM spermidine, 10% glycerol, 1 mM DTT, 0.5 mM PMSF, and 67 mM KCl) at 0°C for 15 min. The ³²P-labeled oligodeoxynucleotide probe (0.2–0.5 ng; 1 μ l) was then added and incubated for 10 min at room temperature. Ficoll dye (20% Ficoll, 0.1% Bromphenol blue) was added to 2%, and the entire reaction was loaded directly onto a 6%/0.5 $\times$ TBE/1.5-mm native polyacrylamide gel. The sequence of the wild-type TCR α enhancer oligodeoxynucleotide was 5'-GATCTAGGGCACCC-TTTGAAGCTCT-3', and the sequence of the double point mutant was 5'-GATCTAGGGCACCCATAGAAGCTCT-3'.

Southwestern experiments were performed as described (Waterman and Jones 1990) except that hybridization solutions contained 4.5 μ g/ml of salmon sperm DNA. Oligodeoxynucleotide probes end-labeled with ³²P were identical to those described for mobility-shift experiments.

In vitro protein synthesis

Twenty micrograms of each of the indicated plasmids was linearized with *Hind*III, phenol-extracted, and precipitated with

ethanol. RNA was synthesized by using T7 polymerase and 5 μ g of the linear templates in 50- μ l reactions that were incubated for 1 hr at 37°C, according to the manufacturer's suggestions (Promega). Reactions were terminated by RQ1 DNase treatment (5 units, Promega) for 15 min at 37°C. RNA was phenol-extracted, ethanol-precipitated, and resuspended in 50 μ l of DEPC-treated distilled water. An aliquot of RNA (2 μ l) was translated by using the rabbit reticulocyte lysate, according to the manufacturer's protocol (Promega), with 2.5 μ l (37.5 μ Ci) of [³⁵S]methionine. A total of 5 μ l of each reaction was then analyzed by denaturing SDS-PAGE (10%). After electrophoresis, the gel was soaked in Enlightening (DuPont) for 30 min before drying and exposing to XAR-5 film.

Northern blot analysis

Northern blots of murine RNA from tissues of 4-week-old mice (except the placenta RNA, which was derived from a 17 day-old fetus) were generously provided by Gail Baughman (The Salk Institute). Total RNA was isolated from the indicated tissues or cell lines using the guanidine/phenol method (Chomczynski and Sacchi 1987). Each RNA was mixed with 0.8 μ g of ethidium bromide (to visualize the RNA in subsequent blotting steps) and separated on a 1% formaldehyde-agarose (HEPES) gel at 40 mA for 16 hr. The RNA was blotted to Hybond N membrane by adsorption overnight in 20 $\times$ SSC and fixed to the nylon support by UV irradiation. Blots were then probed with random-primed labeled DNA following 3 hr of prehybridization at 65°C in 50 mM PIPES, 100 mM NaCl, 50 mM sodium phosphate, 1 mM EDTA, and 5% SDS. Hybridization proceeded overnight in the same buffer at 65°C (except for blots with murine RNA samples, which were hybridized at 60°C) at a DNA probe concentration of 1×10^6 cpm/ml. Blots were washed once at room temperature in 100 ml of 5% SDS, 1 $\times$ SSC (0.015 M sodium citrate, 0.15 M NaCl), followed by a 20-min wash at the hybridization temperature (Virca et al. 1990). Northern blots were exposed to film overnight with a screen. RNA markers (0.24- to 9.5-kb RNA ladder; BRL) were loaded alongside samples and visualized by reprobing the blot with random-primed λ DNA.

Cell culture and transient expression experiments

HeLa SL3 cells (a subclone adapted for spinner culture) were maintained in 1 $\times$ Joklik's media supplemented with 5% bovine calf serum. For transient expression experiments, luciferase plasmids were cotransfected with the indicated TCF-1 α or control expression plasmids using the DEAE dextran-chloroquine method described previously (Waterman and Jones 1990).

Acknowledgments

We thank Connie Hansen for assistance with construction of the λ ZAP II cDNA library, Ralf Schoepfer for advice on various aspects of the cloning procedure, Gail Baughman for sharing murine Northern blots, Harry Mangalam for technical advice and reagents for bacterial expression of TCF-1 α , and both Harry Mangalam and Steven Koerber for assistance in searching data banks. We also thank Christine Schar and Melissa Weiss for excellent technical assistance and members of the Jones lab for their advice during the course of these experiments. This research was funded by grants from the National Institutes of Health, the Mathers Foundation, and the California Universitywide State Task Force on AIDS. W.H.F. is a Clayton Foundation Investigator supported by the Clayton Foundation for Re-

search, California Division. M.L.W. is supported by a postdoctoral fellowship from the American Cancer Society, and K.J. is a member of the Pew Biomedical Scholars Program.

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.

References

- Alt, F.W., T.K. Blackwell, and G.D. Yancopoulos. 1987. Development of the primary antibody repertoire. *Science* **238**: 1079–1087.
- Berta, P., J.R. Hawkins, A.H. Sinclair, A. Taylor, B.L. Griffiths, P.N. Goodfellow, and M. Fellous. 1990. Genetic evidence equating SRY and the testis-determining factor. *Nature* **348**: 448–450.
- Bhat, N.K., K.L. Komschlies, S. Fujiwara, R.J. Fisher, B.J. Mathieson, T.A. Gregorio, H.A. Young, J.W. Kasik, K. Ozato, and T.S. Papas. 1989. Expression of *ets* genes in mouse thymocyte subsets and T cells. *J. Immunol.* **142**: 672–678.
- Bosselut, R., J.F. Duvall, A. Gegonne, M. Bailly, A. Hemar, J. Brady, and J. Grysdael. 1990. The product of the *c-ets-1* proto-oncogene and the related *Ets2* protein act as transcriptional activators of the long terminal repeat of human T cell leukemia virus HTLV-1. *EMBO J.* **9**: 3137–3144.
- Burri, M., Y. Tromvoukis, D. Bopp, G. Frigerio, and M. Noll. 1989. Conservation of the paired domain in metazoans and its structure in three isolated human genes. *EMBO J.* **8**: 1183–1190.
- Chen, J.H. 1985. The proto-oncogene *c-ets* is preferentially expressed in lymphoid cells. *Mol. Cell. Biol.* **5**: 2993–3000.
- Chien, Y., M. Iwashima, K.B. Kaplan, J.F. Elliot, and M.M. Davis. 1987. A new T-cell receptor gene located within the alpha locus and expressed early in T-cell differentiation. *Nature* **327**: 677–682.
- Chomczynski, P. and N. Sacchi. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**: 156–159.
- Davis, M.M. and P.J. Bjorkman. 1988. T-cell antigen receptor genes and T-cell recognition. *Nature* **334**: 395–401.
- Erikson, J., L. Finger, L. Sun, A. Ar-Rushdi, K. Nishikura, J. Minowada, J. Finan, E.S. Emanuel, P.C. Nowell, and C.M. Croce. 1986. Deregulation of *c-myc* by translocation of the α -locus of the T-cell receptor in T-cell leukemias. *Science* **232**: 884–886.
- Fischer, W.H., D. Karr, B. Jackson, M. Park, and W. Vale. 1991. Micro sequence analysis of proteins purified by gel electrophoresis. *Methods Neurosci.* **6**: 69–84.
- Frohman, M.A., M.K. Dush, and G.R. Martin. 1988. Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci.* **85**: 8998–9002.
- Furley, A.J., S. Mizutani, K. Weilbaecher, H.S. Dhaliwal, A.M. Ford, L.C. Chan, H.V. Molgaard, B. Toyonaga, T. Mak, P. van den Elsen, D. Gold, C. Terhorst, and M.F. Greaves. 1986. Developmentally regulated rearrangement and expression of genes encoding the T cell receptor-T3 complex. *Cell* **46**: 75–87.
- 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.
- Havran, W.L. and J.P. Allison. 1988. Developmentally ordered

- appearance of thymocytes expressing different T cell antigen receptors. *Nature* **335**: 443–445.
- Ho, I.-C., L.-H. Yang, G. Morle, and J.M. Leiden. 1989. A T-cell-specific transcriptional enhancer element 3' of C α 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.
- Ho, I.-C. and J.M. Leiden. 1990. The T α 2 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.
- Jantzen, H.-M., A. Admon, S.P. Bell, and R. Tjian. 1990. Nuclear transcriptional factor hUBF contains a DNA-binding motif with homology to HMG proteins. *Nature* **344**: 830–836.
- Johnson, P.F. and S.L. McKnight. 1989. Eukaryotic transcriptional regulatory proteins. *Annu. Rev. Biochem.* **58**: 799–839.
- Karim, F.D., L.D. Urness, C.S. Thummel, M.J. Klemsz, S.R. McKercher, A. Celada, C. Van Beveran, R.A. Maki, C.V. Gunther, J.A. Nye, and B.J. Graves. 1990. The ETS-domain: A new DNA-binding motif that recognizes a purine-rich core DNA sequence. *Genes & Dev.* **4**: 1451–1453.
- 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.
- Klemsz, M.J., S.R. McKercher, A. Celada, C. Van Beveran, 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.
- Kozak, M. 1990. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc. Natl. Acad. Sci.* **87**: 8301–8305.
- 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.
- Lathe, R. 1985. Synthetic oligonucleotide probes deduced from amino acid sequence data: Theoretical and practical considerations. *J. Mol. Biol.* **183**: 1–12.
- Learned, R. M., T.K. Learned, M.M. Haltiner, and R.T. Tjian. 1986. Human rRNA transcription is modulated by the coordinate binding of two factors to an upstream control element. *Cell* **45**: 847–857.
- Marrack, P. and J. Kappler. 1987. The T cell receptor. *Science* **238**: 1073–1079.
- Martinez-Valdez, H., E. Thompson, and A. Cohen. 1988. Coordinate transcriptional regulation of α and δ chains of the T-cell antigen receptors by phorbol esters and cyclic adenosine 5'-monophosphate in human thymocytes. *J. Biol. Chem.* **263**: 9561–9564.
- McDougall, S., C.L. Peterson, and K. Calame. 1988. A transcriptional enhancer 3' of C β 2 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.
- Orkin, S.H. 1990. Globin gene regulation and switching: Circa 1990. *Cell* **63**: 665–672.
- Paillard, F., G. Sterkers, and C. Vaquero. 1990. Transcriptional and post-transcriptional regulation of TcR, CD4 and CD8 gene expression during activation of normal T lymphocytes. *EMBO J.* **9**: 1867–1872.
- Pardoll, D.M., B.J. Fowlkes, J.A. Bluestone, A. Kruisbeck, W.L. Maloy, J.E. Coligan, and R.D. Schwartz. 1987. T cell receptors during thymocyte development. *Nature* **326**: 79–81.
- Redondo, J.M., S. Hata, C. Brocklehurst, and M.S. Krangel. 1990. A T cell-specific transcriptional enhancer within the human T cell receptor δ locus. *Science* **247**: 1225–1229.
- Shackelford, D.A., A.V. Smith, and I.S. Trowbridge. 1987. Changes in gene expression induced by a phorbol diester: Expression of IL-2 receptor, T3, and T cell antigen receptor. *J. Immunol.* **138**: 613–619.
- Sinclair, A.H., P. Berta, M.S. Palmer, J. Ross 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.
- Strominger, J.L. 1989. Developmental biology of T cell receptors. *Science* **244**: 943–950.
- 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.
- 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 HMA box. *EMBO J.* **10**: 123–132.
- Virca, G.D., W. Northemann, B.R. Shiels, G. Widera, and S. Broome. 1990. Simplified Northern blot hybridization using 5% sodium dodecyl sulfate. *Biotechniques* **8**: 370–371.
- Wasylyk, B., C. Wasylyk, P. Flores, A. Begue, D. Leprince, and D. Stehelin. 1990. The *c-ets* proto-oncogenes encode transcription factors that cooperate with c-Fos and c-Jun for transcriptional activation. *Nature* **346**: 191–193.
- Waterman, M.L. and K.A. Jones. 1990. Purification of TCF-1 α , a T cell-specific transcription factor that activates the human T cell receptor C α -gene enhancer in a context-dependent manner. *New Biol.* **2**: 621–636.
- Wilkinson, M.F. and C.L. MacLeod. 1988. Induction of T cell receptor- α and - β mRNA in SL12 cells can occur by transcriptional and post-transcriptional mechanisms. *EMBO J.* **7**: 101–109.
- 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. $\alpha\beta$ lineage-specific expression of the α T cell receptor gene by nearby silencers. *Cell* **59**: 649–655.
- Yamamoto, M., L. Ko, M. Leonard, H. Beug, S. Orkin, and J. Engel. 1990. Activity and tissue-specific expression of the transcription factor NF-E1 multigene family. *Genes & Dev.* **4**: 1650–1662.

A thymus-specific member of the HMG protein family regulates the human T cell receptor C alpha enhancer.

M L Waterman, W H Fischer and K A Jones

Genes Dev. 1991, **5**:

Access the most recent version at doi:[10.1101/gad.5.4.656](https://doi.org/10.1101/gad.5.4.656)

References

This article cites 47 articles, 18 of which can be accessed free at:
<http://genesdev.cshlp.org/content/5/4/656.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).