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

By Mariëtte Oosterwegel,* Marc van de Wetering,* Dennis Dooijes,* Leo Klomp,‡ Astar Winoto,‖ Katia Georgopoulos,¶ Frits Meijlink,§ and Hans Clevers*

*From the *Department of Clinical Immunology, University Hospital Utrecht, 3508 GA, Utrecht; the ‡Department of Cell Biology, Utrecht University, Utrecht; §Hubrecht Laboratory, Utrecht, The Netherlands; the ‖Department of Molecular and Cell Biology, Division of Immunology, University of California Berkeley, Berkeley, California 94720; and the ¶Department of Molecular Biology, Massachusetts General Hospital, Harvard University, Boston, Massachusetts 02138*

Summary

CD3-ϵ gene expression is confined to the T cell lineage. We have recently identified and cloned a human transcription factor, TCF-1, that binds to a functional element in the T lymphocyte–specific enhancer of CD3-ϵ. In a panel of human cell lines, TCF-1 expression was restricted to T lineage cells. TCF-1 belonged to a novel family of genes that contain the so-called high mobility group 1 (HMG) box. Here we report the cloning of murine TCF-1. Two splice alternatives were identified that were not previously observed in human TCF-1. Murine and human TCF-1 displayed a 95.5% overall amino acid homology. Recombinant murine and human TCF-1 recognized the same sequence motif in the CD3-ϵ enhancer as judged by gel retardation and methylation interference assays. With the murine cDNA clones several aspects of TCF-1 were analyzed. First, deletion analysis revealed that a region of TCF-1 containing the HMG box was sufficient for sequence-specific binding. Second, by high stringency Northern blotting and in situ hybridization, TCF-1 expression was shown to be confined to the thymus and to the T cell areas of the spleen. Third, TCF-1 bound specifically to a functional T cell–specific element in the T cell receptor α (TCR-α) enhancer. The T lineage–specific expression and the affinity for functional motifs in the TCR-α and CD3-ϵ enhancers imply an important role for TCF-1 in the establishment of the mature T cell phenotype.

The phenotype of a cell is determined by the complement of actively transcribed structural genes in that cell. A differentiating cell proceeds through a program of phenotype changes. Unique sets of structural genes are expressed at each step. Control of the expression of such sets of genes is exerted principally by developmentally active transcription factors; the activation and inactivation of these transcription factors is believed to be central to the control of phenotype during differentiation (1–3).

In recent years, it has become feasible to identify the *cis*-acting elements (promoters, enhancers, silencers) that control transcription of structural genes expressed at particular stages of differentiation. These *cis*-acting elements can be used to identify DNA-binding transcription factors that recognize specific sequence motifs within these elements. Of special interest are transcription factors that are uniquely active in the cell type where the structural gene is expressed. Several so-called "tissue-specific" transcription factors have thus been identified and cloned. These include the B lymphocyte factor Oct-2, which binds the octamer motif (ATTTGCAT) of the B cell–specific regulatory elements of the Ig genes (4, 5); the pituitary factor Pit-1, which controls growth hormone and prolactin gene expression in somatotroph and lactotroph cells respectively (6, 7); the erythrocyte factor EryFI, which binds to hemoglobin regulatory sequences (8, 9); and MyoD (and several related genes), which can confer the muscle cell phenotype onto fibroblasts (10, 11).

The recent identification of a number of T lymphocyte–specific enhancers (reviewed in reference 12) allows the characterization of transcription factors that govern the differentiation pathway of the T lymphocyte. We have set out to analyze the T cell–specific enhancer of the gene encoding the CD3-ϵ chain of the TCR/CD3 complex (13–15). We have recently described and cloned a transcription factor from human T cells, TCF-1, that is involved in the activity of this enhancer (16). TCF-1 was initially detected by gel

 J. Exp. Med. © The Rockefeller University Press • 0022-1007/91/05/1133/10 $2.00
Volume 173 May 1991 1133–1142

retardation analysis in a screen for T cell–specific proteins binding to the human CD3-ϵ enhancer. Methylation interference footprinting revealed that the TCF-1 cognate motif in the CD3-ϵ enhancer consisted of AAcAAAG (where contact bases are capitalized). TCF-1 was subsequently cloned and was found to be a member of a family of genes with homology to high mobility group I (HMG)[1] proteins. This family includes the polymerase I transcription factor UBF (17), the putative mammalian sex-determining gene SRY (18), and *S. pombe* and *N. crassa* mating type genes (19, 20). Three alternative splice forms of the TCF-1 gene were identified that differed at the extreme COOH terminus of the encoded protein. These alternative splice forms were termed TCF-1A, TCF-1B, and TCF-1C. Within a panel of cell lines, the TCF-1 gene was expressed uniquely in T cells. Finally, recombinant TCF-1 could transactivate transcription through its cognate motif. Here, we report the isolation and analysis of murine TCF-1 cDNA clones.

Materials and Methods

Cells. Human T cell lines Jurkat, Molt4, and Peer; the murine TCR-α/β T cell line EL-4; human B lineage cell lines BS 4.9 and Daudi; and the murine B lineage line Ag-8 were all grown in RPMI 1640 supplemented with 5% FCS and antibiotics. The human cervical carcinoma cell line HeLa and murine fibroblasts NIH-3T3 were grown in DMEM supplemented with 5% FCS and antibiotics.

Nuclear Extracts. Nuclear extracts were prepared by gentle lysis of 10^7–10^8 cells in lysis buffer (30% sucrose [wt/vol]; 40 mM Tris [pH 7.5]; 37 mM KCl; 12 mM $MgCl_2$) in the presence of 0.8% Triton X-100. After two washes with lysis buffer, the nuclei were extracted with 2.5 pellet volumes of extraction buffer (10 mM Hepes [pH 7.9]; 400 mM NaCl; 1.5 mM $MgCl_2$; 0.2 mM EGTA; 20% glycerol) for 30 min on ice. Nuclear debris was removed by centrifugation (15,000 rpm; 5 min). Protein concentration of the clear supernatant was determined and nuclear extracts were stored at −70°C.

Gel Retardation Assay. Annealed oligonucleotides were labeled by T4 polynucleotide kinase and γ-[^{32}P]-ATP. All probes were purified by polyacrylamide electrophoresis. For a typical binding reaction, nuclear extract (5 μg protein) and 1 μg poly dI-dC were incubated in a volume of 15 μl containing 10 mM Hepes, 60 mM KCl, 1 mM EDTA, 1 mM DTT, and 12% glycerol. After 5 min of preincubation at room temperature, a probe (20,000 cpm) was added and the mixture was incubated for an additional 20 min. The samples were then subjected to electrophoresis through a nondenaturing 4% polyacrylamide gel run in 0.25 * TBE at room temperature. In competition experiments, nonlabeled competitor DNA was added with the poly dI-dC.

Probes Used. MWϵ1 (identical to MW56 in reference 16): GGGAGACTGAGAACAAAGCGCTCTCACAC annealed to CCCGTGTGAGAGCGCTTTGTTCTCAGTCT. MWϵ1Sac: GGGAGACTGAGCCGCGGTCGCTCTCACAC annealed to CCCGTGTGAGAGCGACCGCGGCTCAGTCT. MWϵ1A1G: GGGAGACTGGGAACAAAGCGCTCTCACAC annealed to CCCGTGTGAGAGCGCTTTGTTCCCAGTCT. MWδB: GGGTAATTAAAAATAAAGAACATGAT annealed to CCCATCATGTTCTTTATTTTTAATTA. MWpal: GGGCCCCCTCTGCAAGCAGAGTGT annealed to GGGACACTCTGCTTGCAGAGGGGG. MWTCF-1α: GGGCAAGTAGGGCACCCTTTGAAGCTCT annealed to CCCAGAGCTTCAAAGGGTGCCCTACTTG. All oligonucleotides were synthesized on a cell synthesizer (381A; Applied Biosystems, Inc., Foster City, CA).

Isolation of cDNA Clones. Murine TCF-1 cDNA clones were isolated from an EL-4 cDNA library in λ-ZAP (10^6 primary recombinant phages; average insert size, 1.2 kb) by standard hybridization screening at low stringency using the insert of ΦTCF-1 (16) as the probe (21).

DNA Sequencing. Sequencing was performed on double-stranded DNA templates using T7 DNA polymerase according to the manufacturer's instructions (Pharmacia Fine Chemicals, Piscataway, NJ). Sequencing templates were generated by subcloning the inserts of isolated cDNA clones, and BA131 deletions thereof into pUC19. Sequencing primers were the universal and reverse M13 primers, and oligonucleotides representing specific internal sequences of the cDNA clones.

RNA Isolation and Analysis. Total RNA was prepared by cell lysis in RNAzol according to the manufacturer's instructions (Cinna-Biotecx), followed by phenol-chloroform extraction and 2-propanol precipitation. 10 μg of total RNA per lane was run for Northern analysis on 1% agarose containing 6% formaldehyde. RNA was transferred to nitrocellulose and hybridized with murine TCF-1 cDNA probes labeled by random oligo priming, all according to standard procedures (21).

In Situ Hybridization. In situ hybridization on paraffin-embedded thymus and spleen sections (2-wk-old male BALB/c) was performed essentially as described by Wilkinson et al. (22), with modifications introduced by Kress et al. (23). An antisense ^{35}S-labeled RNA TCF-1 probe was generated using T7 RNA polymerase and ^{35}S-UTP. Template was prepared by subcloning the insert of pM2a into the EcoR1 site of pBluescript SK (Stratagene); the resulting construct was digested with NsiI (site at bp 941) to yield an ~750-bp uniformly labeled probe covering 3′ untranslated and 65 bp of coding sequence of alternative splice form B. This probe did not cross-hybridize to sequences other than TCF-1 in the human and mouse genome under conditions of reduced stringency (data not shown). Furthermore, as predicted from the structure of the human TCF-1 gene, this probe represented sequences present in all alternative TCF-1 mRNA species. The probe size was reduced to 200–300 bp by limited hydrolysis in NaOH.

Expression of Murine TCF-1 in Escherichia coli. *E. coli* strain N4830-1 was used as the host for expression of TCF-1 using the heat-inducible protein A expression vector pRIT2-T (24). 100 ml of LB containing 100 μg/ml ampicillin was inoculated with 1 ml of an overnight culture, and grown at 30°C to A_{600} ~0.9. One volume LB of 54°C was added and the cells were grown at 42°C for 90 min. Cells were collected by centrifugation at 5,000 rpm for 10 min, and the pellet was resuspended in 5 ml of PBS containing 1 mM PMSF. The bacteria were lysed by sonification (four times for 4 min) on ice. Cell fragments were removed by centrifugation (10 min at 15,000 *g*). The fusion protein was purified over IgG-Sepharose according to the manufacturer's instructions (Pharmacia Fine Chemicals) and stored at −70°C. All fusion protein batches were analyzed by SDS/polyacrylamide gel electrophoresis.

Constructs used were: (*a*) pRIT-M2a, generated from pRIT-TCF-1A (16) by digestion with PstI and insertion of the PstI fragment of pM2a (bp 586–1121). (*b*) 3′ deletion clones were generated by opening pRIT-M2a at the NsiI site (bp 941), followed by limited digestion with Bal31 exonuclease, and insertion of an oligonucleotide linker containing stop codons in all frames. (*c*) 5′ deletions were generated by insertion of the PstI fragment of pM2a (bp 586–1121) into pRIT2-T (24), resulting in a 5′ truncation, starting

[1] *Abbreviation used in this paper:* HMG, high mobility group.

at amino acid 133); and insertion of a fragment, generated from pM2a by PCR, into SmaI/PstI-digested pRIT2-T, resulting in a truncation starting at amino acid 155. All constructs were sequenced to determine the deletion start/end points and to exclude PCR errors.

Methylation Interference Footprinting. MWε1 was labeled at the positive strand oligonucleotide with γ-[^{32}P]-ATP using T4 polynucleotide kinase. After annealing, the probe was purified over polyacrylamide. The labeled probe was partially methylated at purine residues using dimethylsulfate. 100,000 cpm of methylated probe was used in a scale-up of the gel retardation binding reaction. After fractionation by gel retardation assay, the wet gel was subjected to autoradiography. The bound and free probes were cut out and recovered by electroelution. After cleavage by NaOH at G and A residues, the sequence was analyzed on a 10% polyacrylamide–8 M urea sequencing gel.

Results

Gel Retardation Analysis Indicates the Existence of a Murine TCF-1 Homologue. To demonstrate the presence of a TCF-1-like binding activity in nuclear extracts from murine T cells, we performed gel retardation analysis with probe MWε1 (which contains the cognate motif of human TCF-1). Analysis of nuclear extracts derived from a panel of murine cells resulted in the appearance of a retarded band that displayed the same sequence preference and T cell specificity as was observed with TCF-1 in human T cell extracts. Fig. 1 depicts typical results obtained with extracts from the murine thymoma cell line El-4, the murine B lineage cell line Ag-8, and murine NIH-3T3 fibroblasts, as compared with extracts of several human T and non-T cell lines. The TCF-1 retardation band is indicated with a solid arrow. No TCF-1-like binding was observed with probe MWε1Sac, in which the AAcAAAG motif is replaced by CCGCGGT (data not shown).

Cloning of Murine TCF-1. A λ-ZAP cDNA library was then prepared from the murine TCR/CD3$^+$ EL4 thymoma cell line (10^6 primary recombinant phages, average insert size, 1.2 kb) and amplified. 10^6 plaques were plated out and screened with the radiolabeled insert of ΦTCF-1, the original human TCF-1 bacteriophage clone (16). 10 positive plaques were isolated. The phages could be divided into two groups based on restriction digests of their inserts. The inserts of two representative phages, ΦM2a and ΦM5, were subcloned in pUC19, and the resulting plasmids (pM2a and pM5) were sequenced.

Fig. 2 *A* depicts the sequence of the insert of pM2a. By comparison with the human TCF-1 sequence, pM2a appeared to encode a full-length version of the TCF-1 protein, with a COOH terminus of the TCF-1B splice form (16). Overall sequence homology with human TCF-1B was 95.5% at the amino acid level (see Fig. 2 *C*). An in-frame stop codon (bp 13) was present upstream from the first in-frame ATG (bp 190) and is indicated by a solid bar. Comparison of the sequence of pM2a and pM5 revealed the presence of an extra stretch of 93 bp in pM2a starting at position 489 (Fig. 2 *A*). This stretch was inserted into the proline-rich domain of TCF-1 and itself encoded a proline-rich amino acid sequence (see

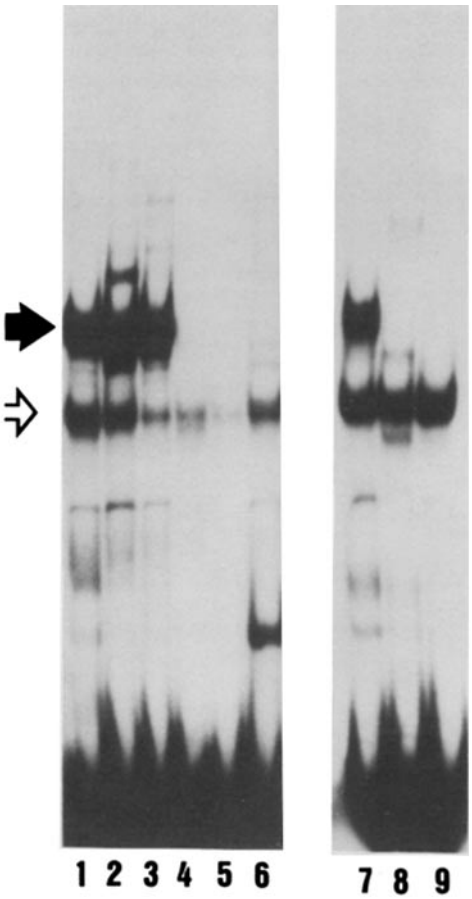


Figure 1. Identification of a TCF-1-like binding activity in murine T cells. Gel retardation analysis was performed with the TCF-1 cognate probe MWε1 in conjunction with nuclear extracts of various human and murine T and non-T cells. The solid arrow indicates the position of TCF-1 from human and murine T cells. The open arrow indicates the position of another (ubiquitous) DNA-binding protein unrelated to TCF-1. Extracts were prepared from: human cell lines (*left*): lane *1*, Jurkat; lane *2*, Molt4; lane *3*, Peer; lane *4*, BS 4.9; lane *5*, Daudi; lane *6*, HeLa. Murine cell lines (*right*): lane 7, El-4; lane *8*, Ag-8; lane *9*, NIH-3T3. Lanes *1–3* and *7* are T lineage cells; lanes *4*, *5*, and *8* are B lineage cells. HeLa is a cervical carcinoma cell line. NIH-3T3 is a murine fibroblast cell line.

Fig. 2 *B*: proline-rich alternative splice). The site of insertion, marked with *P* in Fig. 2 *A*, coincided with exon-intron boundaries in the human TCF-1 gene (M. van de Wetering, F. Holstege, and H. Clevers, manuscript in preparation), indicating that pM2a represented a genuine alternative splice; it was not previously encountered among 34 independent human TCF-1 cDNA clones (16). From Northern blot analysis, the principal murine TCF-1 mRNA was estimated to be ~2,900 bp (see below). Thus, pM2a lacked some 1,200 bp of 3′ untranslated sequence.

The shorter clone M5 represented a splice alternative, which was termed TCF-1D and which we recently also isolated from human T cells (M.W. and H. Clevers, unpublished results). As compared with M2a, 23 bp were inserted near the COOH terminus of the encoded protein (marked with *D* in Fig. 2 *A*). The sequence of this stretch is given in Fig. 2 *B*. The presence of a stop codon at bp 21 of the *D* insert predicted

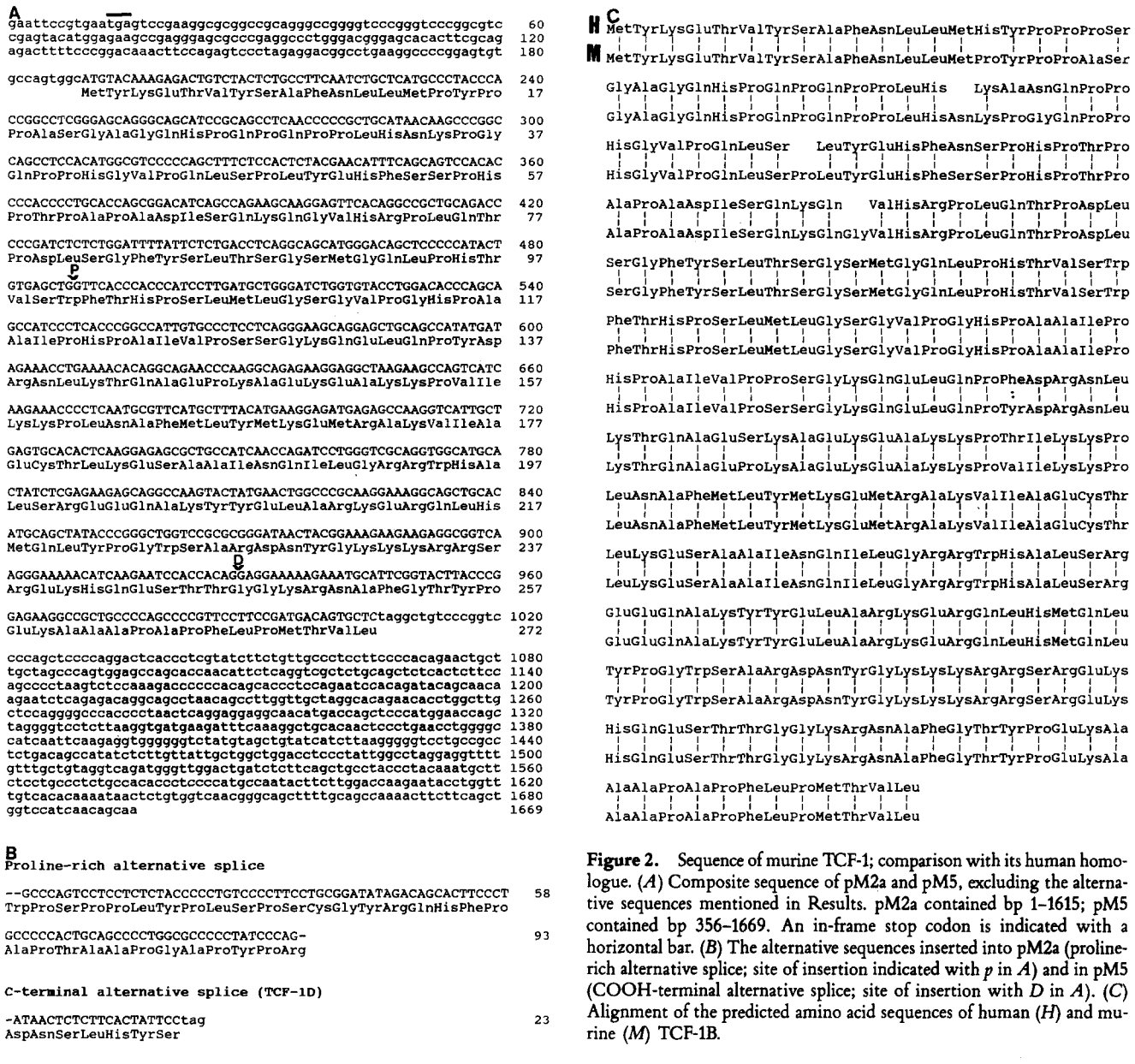


Figure 2. Sequence of murine TCF-1; comparison with its human homologue. (*A*) Composite sequence of pM2a and pM5, excluding the alternative sequences mentioned in Results. pM2a contained bp 1–1615; pM5 contained bp 356–1669. An in-frame stop codon is indicated with a horizontal bar. (*B*) The alternative sequences inserted into pM2a (proline-rich alternative splice; site of insertion indicated with *p* in *A*) and in pM5 (COOH-terminal alternative splice; site of insertion with *D* in *A*). (*C*) Alignment of the predicted amino acid sequences of human (*H*) and murine (*M*) TCF-1B.

a truncated version of the TCF-1 protein. The site of insertion of the *D* segment coincided with the position where human TCF-1A/B/C/D alternative splicing occurs. In the human gene, TCF-1D results from the use of an alternative splice acceptor site located within the last exon (M. van de Wetering et al., manuscript in preparation).

Recombinant Murine TCF-1 Recognizes the AAcAAAG Motif. To investigate the binding specificity of murine TCF-1, we generated a Staphylococcal protein A/TCF-1 fusion protein using the vector pRIT2-T in the appropriate *E. coli* host (24). The fusion protein was purified to near homogeneity by a single IgG-Sepharose chromatography step. Gel retardation analysis performed with the purified murine TCF-1B fusion protein demonstrated specific binding to the MWε1 probe (see Fig. 3 *A*); no binding occurred with the MWε1Sac control probe. Subsequent methylation interference footprinting revealed that the positive strand contact bases of recombinant murine TCF-1 were identical to those of human TCF-1 (indicated by a solid bar in Fig. 3 *B*). Under the experimental conditions, contacts with both A and G appear as spots of decreased intensity.

The HMG Box of TCF-1 Is Involved in Sequence-specific Binding. To delineate the domain in TCF-1 that mediates recognition of the AAcAAAG motif, a set of deletion clones of murine TCF-1 was generated in the pRIT2-T vector by Bal31 and PCR techniques. Fusion proteins were purified and subjected to gel retardation analysis using the cognate probe MWε1 and the negative control probe MWε1Sac. Fig. 4 summarizes the results obtained with these deletion clones. DNA binding was confined to the region from amino acid 133 to

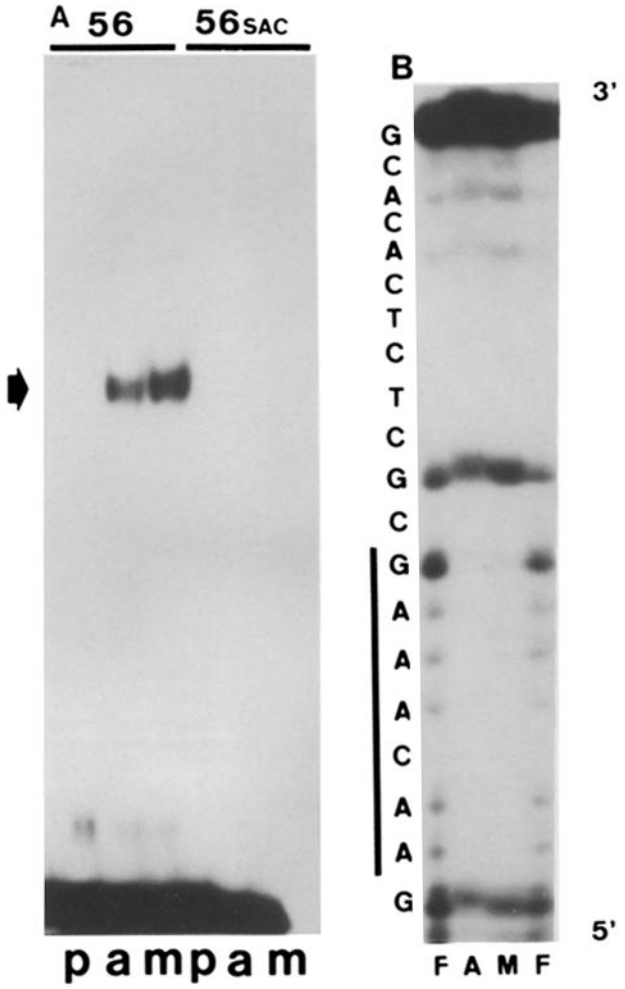


Figure 3. Recombinant murine TCF-1 binds to the TCF-1 motif in the CD3-ϵ enhancer. (*A*) Gel retardation analysis performed with recombinant murine TCF-1B (*M*) and human TCF-1A (*A*) resulted in a specific retardation band with probe MWϵ1 (*left*, indicated with 56), but not with the control probe MWϵ1Sac (*right*, indicated with 56Sac). No retarded bands were observed with control recombinant protein A (*P*). (*B*) Positive strand methylation interference footprinting was performed with probe MWϵ1 in conjunction with recombinant murine TCF-1B (*M*) and human TCF-1A (*A*). Essentially identical footprints were obtained for both proteins (indicated with a bar, 5'AAcAAAG3') in comparison with the reaction performed on the free probe (*F*). Due to the paucity of purines in the relevant area of the negative strand of MWϵ1, no contact bases could be defined by methylation interference on this strand (16).

237. This region encompassed the 77–amino acid HMG box in TCF-1 (amino acids 157–233) as delineated by sequence comparison with other members of the HMG box family (16). It should be noted that a 5′ deletion clone starting at the NH_2-terminal boundary (amino acid 155) of the HMG box did not bind DNA. It thus appeared that residues directly NH_2 terminal to the HMG box (but still located within the TCF-1 basic domain; see Fig. 4) were indispensable for cognate interaction with the TCF-1 DNA motif. However, we cannot exclude the possibility that the loss of binding resulted from interference of the protein A fusion segment with an otherwise intact HMG box in this particular deletion mutant.

Expression of TCF-1 Is Confined to the T Cell Lineage. Low stringency hybridization has revealed the presence of several sequences related to TCF-1 in the human and murine genomes (H. Clevers, unpublished results). Crosshybridization to TCF-1-like genes might therefore hamper expression analysis by Northern blotting at decreased stringency. The availability of murine cDNA probes allowed examination of TCF-1 expression in various tissues of the mouse under conditions of high stringency.

RNA was prepared from 2-wk-old BALB/c mice, and Northern blotting was performed using a 3′ untranslated TCF-1 probe. A predominant 2.9-kb mRNA species, comparable with the size of human TCF-1 mRNA, was thus observed (see Fig. 5 *a*). TCF-1 mRNA was detected only in thymic and splenic RNA.

To investigate patterns of TCF-1 expression within the spleen and thymus, we performed in situ hybridization experiments on paraffin-embedded sections of these lymphoid organs. In situ hybridization of spleen with an antisense ^{35}S-labeled probe revealed that TCF-1 expression was limited to the periarteriolar lymphocyte sheaths (PALS), the T cell areas of the spleen. No signal was observed in red pulp, nor in the B cell follicles and marginal zone (See Fig. 5 *b*). In situ hybridization of thymus sections with the same probe demonstrated that TCF-1 was expressed throughout the thymus, with particularly high levels in the cortex (Fig. 5 *b*). In the same experiment, no signal was obtained with the nonrelevant negative control probe S8 (F. Meijlink, unpublished results) on thymus and spleen sections (data not shown).

TCF-1 Binds to a Functional Motif in the TCR-α Enhancer. Our experiments have demonstrated that TCF-1 controls the T cell–specific activity of the CD3-ϵ enhancer (16; M. Oosterwegel and H. Clevers, unpublished results). If TCF-1 were to play a more general role in T cell differentiation, it should control transcription of other T lymphocyte genes. We therefore compared the TCF-1 footprint with functional motifs in other TCR and CD3 regulatory sequences (12). Homologies to sequences within promoters/enhancers not known to bind T cell–specific factors or to mediate T cell–specific transactivation were not included in the comparison. Two motifs were thus selected: the δB element in the CD3-δ enhancer (25), and the NFα3/TCF-1α element in the TCR-α enhancer (26–29). The relevant oligonucleotide sequences are aligned in Fig. 6 *A*. Specific binding of TCF-1 to the selected oligonucleotide probes was analyzed by gel retardation and competition with relevant and nonrelevant oligonucleotide probes, performed both on recombinant TCF-1 and on crude nuclear extracts.

Fig. 6 *B* presents the results of a gel retardation experiment performed with the murine TCF-1B fusion protein and various oligonucleotide probes. Strong binding was observed to the probes containing the TCF-1 and the NFα3/TCF-1α motifs (probes MWϵ1 and MWTCF-1α, respectively; Fig. 6 *B*, lanes *1* and *4*). Slightly decreased binding was obtained with a mutant version of MWϵ1, in which the A residue located 2 bp upstream of the TCR-1 footprint was replaced by a G (MWϵ1A1G; lane *2*). No binding was observed to the mutant version of MWϵ1, which lacked the TCF-1 motif

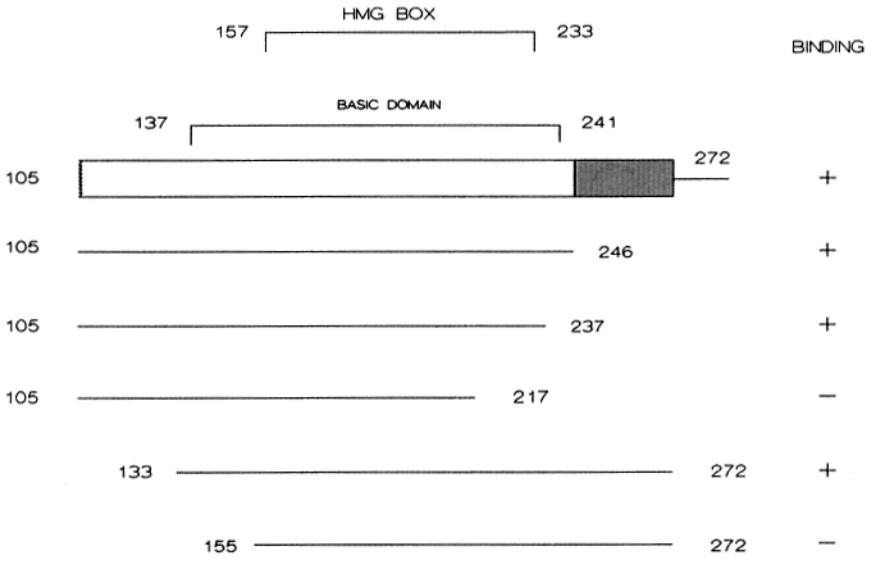


Figure 4. Delineation of the DNA-binding domain of murine TCF-1 by deletion analysis. The indicated deletions of murine TCF-1B were generated as described in Materials and Methods, and subcloned into pRIT2-T to generate protein A fusion proteins. DNA binding was tested by gel retardation with the cognate probe MWϵ1 using purified fusion proteins (see also Fig. 3 *A*). DNA binding was confined to amino acids 133–237, a region encompassing the HMG box.

(MWϵ1Sac; lane *3*). A very weak signal was obtained with the MWδB probe, which contained the δB motif (MWδB; lane *5*). No signal was observed with a probe representing bp 43–65 of the CD3-ϵ enhancer (reference 16; MWpal, lane *6*). The latter sequence has been predicted to interact with TCF-1α by Waterman and Jones (29, see also Fig. 6 *A* and Discussion).

These observations were extended by competition experi-

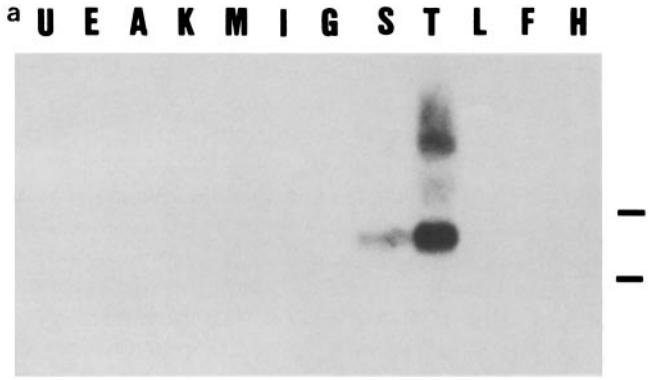
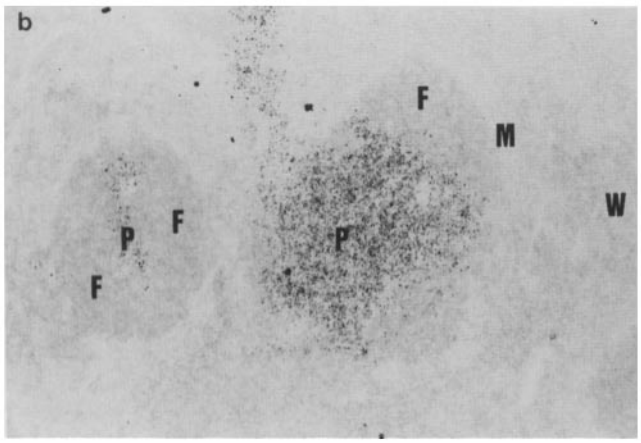
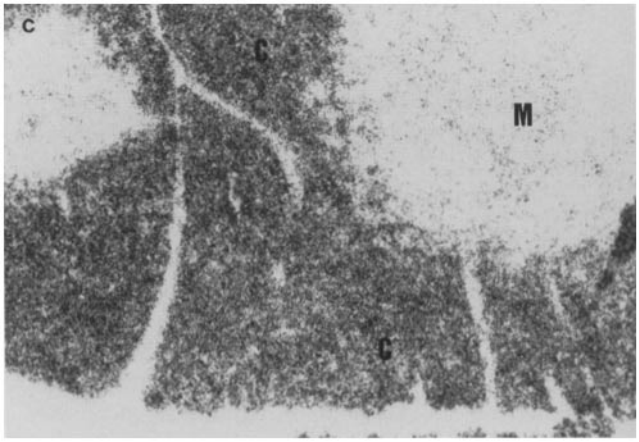


Figure 5. Murine TCF-1 expression is limited to T lineage cells. (*A*) Northern blotting, performed on total RNA prepared from various tissues of 2-wk-old BALB/c mice, only detects TCF-1 mRNA in thymus and spleen samples. The probe used was an NsiI/EcoRI fragment of pM5 (bp 941–1669; see Materials and Methods). The position of the rRNA bands are indicated with horizontal bars. Equal amounts of RNA were present in each lane as judged by examination of the ethidium bromide–stained gel. Tissues: U, uterus; E, testis; A, adrenal gland; K, kidney; M, muscle; I, small intestine; G, stomach; S, spleen; T, thymus; L, liver; F, forebrain; H, hindbrain. The nature of the larger TCF-1 mRNA (estimated to be at least 8 kb) is unknown at present. Prolonged exposure did not reveal expression of TCF-1 mRNA in tissues other than thymus and spleen (data not shown). (*B*) In situ hybridization with a TCF-1 probe of sections of spleen (*left*) and thymus (*right*). Spleen: the autoradiographic signals of labeling were visualized in the periarteriolar lymphocyte sheaths (*P*) of the spleen. No signal was observed in the follicles (*F*), the marginal zone (*M*), and the red pulp (*W*). Thymus: a signal was obtained throughout the thymus but was more prominent in cortex (*C*) than in medulla (*M*).





ments using MWε1 and MWTCF-1α as probes in conjunction with the TCF-1B fusion protein. Fig. 6 *C* (*left*) depicts results obtained with the MWε1 probe. Specific competition was obtained with excess unlabeled MWε1 and MWTCF-1α. No competition was seen with MW56Sac and MWδB. Essentially identical results were obtained in the reciprocal experiment (Fig. 6 *C*, *right*), where cold excess MWε1 and MWTCF-1α specifically competed for binding to the MWTCF-1α probe.

Similar experiments were conducted with crude nuclear extracts of the human T cell line Jurkat. Fig. 6 *D* (*left*) demonstrates that the TCF-1 retarded band (*arrow*) obtained with probe MWε1 could be competed away by cold excess MWε1 and MWTCF-1α, but not by MW56Sac, nor by MWδB. A band with similar migration characteristics was obtained with the MWTCF-1α probe (Fig. 6 *D*, *right*) and probably represented binding of TCF-1. Again, this band was competed away specifically with cold excess MWε1 and with MWTCF-1α.

Taken together, these experiments indicated that TCF-1 bound specifically and with comparable affinities to both the NFα3/TCF-1α probe and to the TCF-1 motif of the CD3-ε enhancer. As summarized in Fig. 6 *A*, no significant binding was observed to the δB element or to various control probes.

Discussion

The present report describes the isolation and analysis of murine TCF-1 cDNA clones. Comparison with human TCF-1 revealed a striking degree of homology, in particular in the COOH-terminal DNA-binding half of this transcription factor. The two murine clones described in detail were homologues of the human alternative splice forms TCF-1B (16) and TCF-1D (this report). The proline-rich alternative exon identified in clone M2a defined a second position in TCF-1 where alternative splicing can occur, and it stressed the significance of the proline residues in the NH_2-terminal half of TCF-1.

We presently do not know what the differences in function are between these alternative splice forms. Based on analogy to CTF/NF-1 (30), the proline-rich NH_2-terminal segment of TCF-1 might constitute a transactivation domain. Extension of the proline-rich domain by an insertional alternative exon could enhance or otherwise modulate transactivation as exerted by TCF-1.

The alternative splices TCF-1A/B/C/D are located near, but (as shown in the present study) COOH terminal to the DNA-binding domain. Indeed, the methylation interference footprints of human TCF-1A, -1B, and -1C were indistinguishable (16). Although the alternative COOH termini are therefore not involved in DNA binding, the complete conservation of human and murine TCF-1B and -1D suggests an important function for these segments. Most likely, the alternative COOH termini provide surface for interaction with other proteins directly or indirectly bound to enhancer sequences.

TCF-1 belongs to a novel family of proteins containing an 80–amino acid motif, which, based on homology to HMG I proteins, has been termed the HMG box (16–20). The only other protein with proven sequence-specific DNA-binding characteristics in this family is the polymerase I transcription factor UBF (17). UBF contains four HMG boxes, which are more or less evenly distributed over the length of the molecule. A deletion mutant of UBF containing the first 204 residues and only one of the HMG boxes was still capable of binding DNA as assessed by binding to a DNA affinity column; the deletion of 104 additional COOH-terminal amino acids removed this HMG box and abrogated DNA binding. From this experiment, the HMG box was proposed to constitute a novel protein motif capable of sequence-specific DNA binding (17).

The presence of a single HMG box in TCF-1 in conjunction with its well-defined recognition motif allowed a direct assessment of the contribution of the TCF-1 HMG box to sequence-specific DNA binding. Our deletion analysis demonstrated that a segment of TCF-1, containing the HMG box plus a short stretch of basic residues directly NH_2 terminal to this box, was sufficient to mediate binding. The future identification of DNA sequence motifs recognized by other members of the HMG box family will allow, by means of sequence comparison and "domain swapping", the identification of critical residues in (and around) the DNA-binding HMG box domain.

Our recognition of a DNA-binding HMG box in a polymerase II transcription factor has important implications for other members in this family, in particular for the candidate mammalian sex-determining gene SRY (18) and for the *S. pombe* and *N. crassa* mating type genes (19, 20). It is likely that these genes, like TCF-1, encode polymerase II transcription factors; they could thus govern sexual differentiation (of mammals, yeast, and fungus, respectively) by direct binding to regulatory sequences of target genes.

We originally identified TCF-1 as a DNA-binding activity present uniquely in nuclear extracts of T lineage cells within a panel of 15 cell lines. The present study extends this observation: Northern blot analysis of mouse tissue RNA showed that TCF-1 mRNA expression was confined to tissues of lymphoid origin. By in situ hybridization, it was subsequently found that TCF-1 was expressed throughout the thymus and in the T cell areas of the spleen. These results indicate that TCF-1 expression is indeed T cell specific, and is initiated early in the T cell differentiation pathway, coinciding with or possibly preceding expression of the CD3 genes (15). We cannot exclude at present that TCF-1 is expressed in other tissues during embryogenesis, or in cells that occur in relatively low numbers in an adult animal. In situ hybridization experiments are currently being conducted to investigate these possibilities.

The expression pattern of TCF-1 is suggestive of its involvement in T cell differentiation. If indeed TCF-1 were a regulator of T cell differentiation, it should control the expression of other genes that build the mature T cell phenotype. As demonstrated here, recombinant TCF-1 was capable of binding to a functional motif in the TCR-α enhancer, which has been termed NFα3 in the murine enhancer (26), and Tα2 (27, 28) and TCF-1α (29) in the human enhancer. These motifs

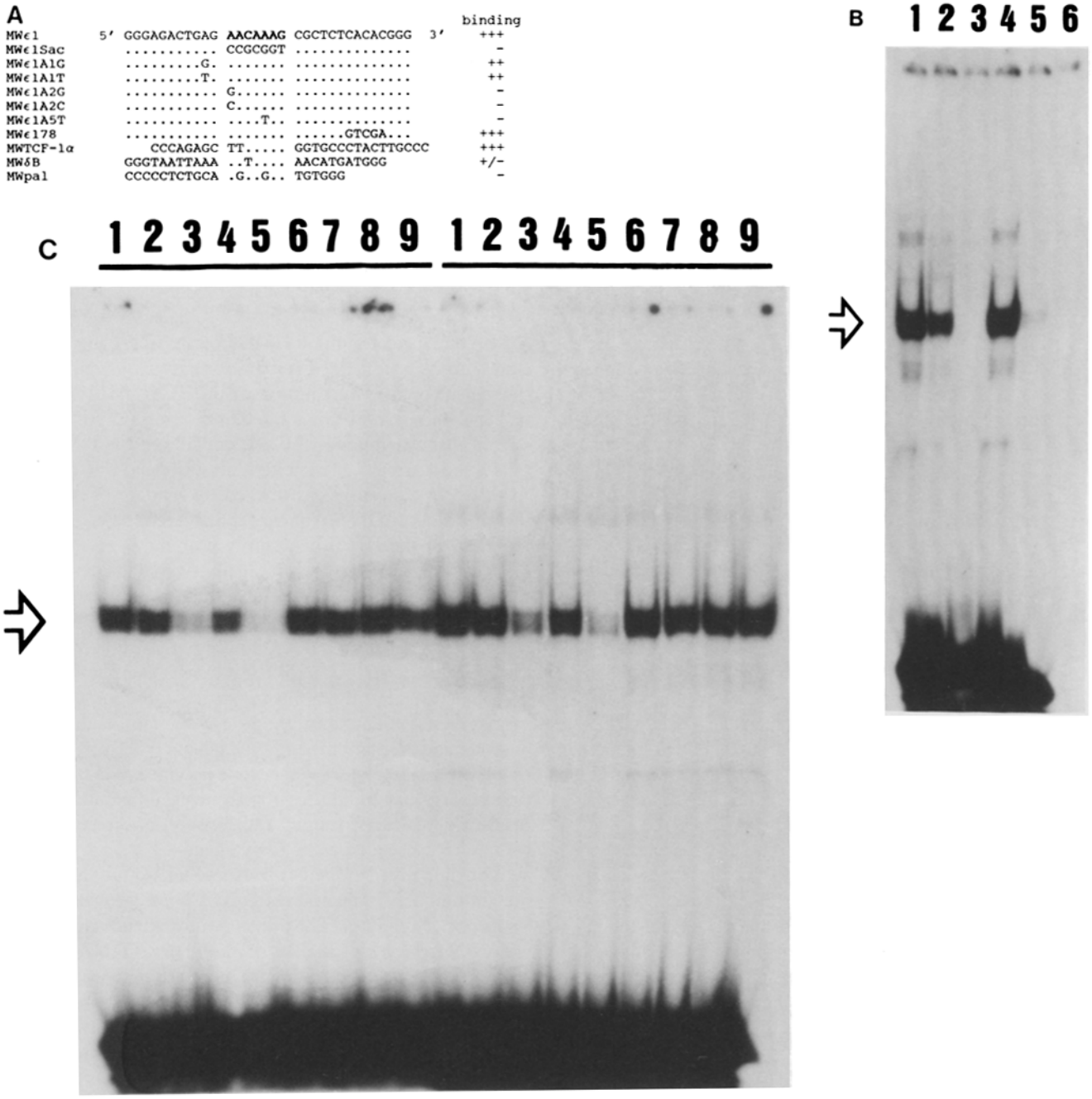
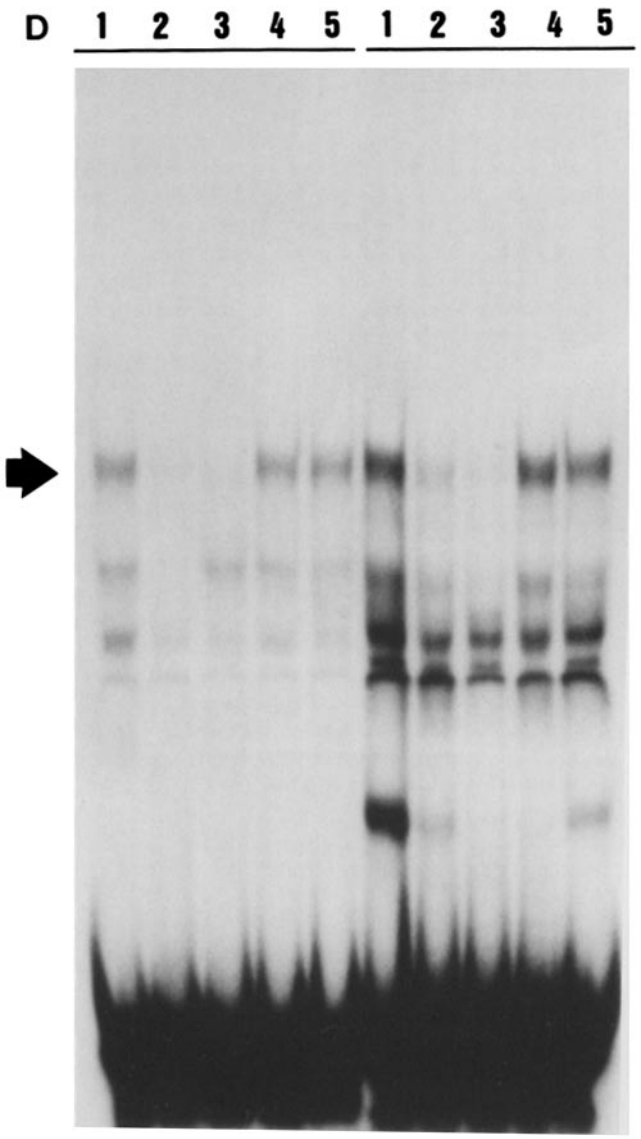


Figure 6. TCF-1 binds specifically to the NFα3/TCF-1α motif of the TCR-α enhancer (26–29). (*A*) Alignment of the sequences of the TCF-1 cognate probe MWε1, and various mutants thereof; of the NFα3/TCF-1α MWTCF-1α (bp 54–81; reference 27); of the δB cognate probe MWδB (bp 361–384 of the CD3-δ enhancer; reference 25); and of MWpal, a sequence in the CD3-ε enhancer (bp 43–65; reference 14) predicted to bind TCF-1α (29). Binding was assessed in the gel retardation assay as described in Results. Results with the point mutants of MWε1 and with MWε178 (M. Oosterwegel and H. Clevers, unpublished results) are not mentioned elsewhere in the text; they are presented to illustrate the sequence specificity of TCF-1. See also *B–D*. (*B*) Gel retardation analysis performed with recombinant murine TCF-1B. Probes are: lane *1*, MWε1; lane *2*, MWε1A1G; lane *3*, MWε1Sac; lane *4*, MWTCF-1α; lane *5*, MWδB; and lane *6*, MWpal. Strong retardation bands (*arrow*) were obtained only with the TCF-1 cognate probes (lanes *1* and *2*) and the NFα3/TCF-1α cognate probe (lane *4*). (*C*) Competition analysis performed with recombinant murine TCF-1B and the probes MWε1 (*left*) and MWTCF-1α (*right*). Competition was performed with: lane *1*, no competition; lanes *2* and *3*, 10- and 100-fold excess MWε1; lanes *4* and *5*, 10- and 100-fold excess MWTCF-1α; lanes *6* and *7*, 10- and 100-fold excess MWε1Sac; lanes *8* and *9*, 10- and 100-fold excess MWδB. Competition was seen only with MWε1 and MWTCF-1α. (*D*) Gel retardation analysis performed with crude nuclear T cell extracts and the probes MWε1 (*left*) and MWTCF-1α (*right*). Competition was performed with: lane *1*, no competition; lane *2*, 100-fold excess MWε1; lane *3*, 100-fold excess MWTCF-1α; lane *4*, 100-fold excess MW56Sac; and lane *5*, 100-fold excess MWδB. The arrow indicates the position of the TCF-1 retardation band. Specific competition occurred again with cold MWε1 and MWTCF-1α.



have been defined by DNaseI footprinting. The NFα3 and TCF-1α DNA-binding factors are T cell specific and appear to be homologues of the same factor. The Tα3 motif probably represents a composite footprint of NFα3/TCF-1α with a second T cell–specific DNA-binding protein (28, 29).

It is highly likely that the TCF-1, as present in T cell nuclear extracts, is responsible for the NFα3/TCF-1α T cell–specific DNaseI footprint (27). Recently, Waterman and Jones (29) have purified a T cell–specific DNA-binding protein, TCF-1α, that binds to the NFα3 motif. By analyzing the affinity of this protein for a panel of double-stranded DNA molecules, they formulated a consensus binding site: CTNTG (or CANAG) for this protein. Although this consensus appears too degenerate to describe the sequence preference of a biologically relevant transcription factor, it fits the TCF-1 motif (AAcAAAG). The cloning of TCF-1α has been announced (29); we anticipate that sequence comparison will reveal TCF-1α to be identical to human TCF-1.

TCF-1 thus appears to be involved in the transcriptional control of at least two T cell genes: CD3-ε and TCR-α. In the analysis of a set of prothymocyte cell lines, TCF-1 expression was strictly concordant with CD3-ε expression (M. Oosterwegel, A. Kruisbeek, and H. Clevers, unpublished results). However, several of the cell lines containing TCF-1 do not express TCR-α. Moreover, the human leukemic T cell line CCRF-CEM expresses TCF-1 and CD3-ε but similarly does not express TCR-α, despite the presence of a productively rearranged TCR-α locus (16, 31). This suggests that TCF-1 expression might be sufficient for CD3-ε expression by differentiating thymocytes, but that TCR-α expression requires the presence of additional stage-specific transcription factors.

We propose the existence of a set of genes controlling T cell differentiation, possibly in a cascade-like fashion. For the expression of individual structural T cell genes, different combinations of these differentiation control genes act in concert with more ubiquitous transcription factors. The combinatorial activity of a few T cell–specific transcription factors, each appearing at a particular stage of T cell differentiation, can thus explain the intricate molecular events that accompany T cell differentiation.

We thank Dr. H.-J. Schuurman for help with the histology; Dr. P. Peters and F. Holstege for critically reading the manuscript; T. van Rijn for photography; A. Postma for expert secretarial help; and the members of the lab for helpful discussions.

This work was supported by the Department of Clinical Immunology, University Hospital Utrecht. K. Georgopoulos is a special fellow of the Leukemia Society of America. H. Clevers is a fellow of NWO.

Address correspondence to Hans Clevers, Department of Clinical Immunology, University Hospital Utrecht, P.O. Box 85500, 3508 GA, Utrecht, The Netherlands.

Received for publication 19 November 1990.

References

1. Maniatis, T., S. Goodbourn, and J.A. Fischer. 1987. Regulation of inducible and tissue-specific gene expression. *Science (Wash. DC).* 236:1237.
2. Atchison, M.L. 1988. Enhancers: mechanisms of action and cell specificity. *Annu. Rev. Cell Biol.* 4:127.
3. Johnson, P.F., and S.L. McKnight. 1989. Eukaryotic transcriptional regulatory proteins. *Annu. Rev. Biochem.* 58:799.
4. Mueller, M.M., S. Ruppert, W. Schaffner, and P. Matthias. 1988. A cloned octamer transcription factor stimulates transcription from lymphoid-specific promoters in non-B cells. *Nature (Lond.).* 336:544.
5. Scheidereit, C., J.A. Cromlish, T. Gerster, K. Kawakami, C.-G. Balmaceda, A. Currie, and R.G. Roeder. 1988. A human lymphoid-specific transcription factor that activates immunoglobulin genes is a homeobox protein. *Nature (Lond.).* 336:551.
6. 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.
7. 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.
8. Evans, T., and G. Felsenfeld. 1989. The erythroid-specific transcription factor EryF1: A new finger protein. *Cell.* 58:877.
9. Tsai, S.-F., D.I. Martin, L.I. Zon, A.D. D'Andrea, C.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 (Lond.).* 339:446.
10. Davis, R.L., H. Weintraub, and A.B. Lassar. 1987. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell.* 51:987.
11. Lassar, A.B., J.N. Buskin, D. Lockshon, R.L. Davis, S. Apone, S.D. Hauschka, and H. Weintraub. 1989. MyoD is a sequence-specific DNA binding protein requiring a region of myc homology to bind to the muscle creatine kinase enhancer. *Cell.* 58:823.
12. Clevers, H., and M. Owen. 1991. Towards a molecular understanding of T cell differentiation. *Immunol. Today.* In press.
13. Clevers, H.C., S. Dunlap, T.E. Wileman, and C. Terhorst. 1988. Human CD3-ϵ gene contains three miniexons and is transcribed from a non-TATA promoter. *Proc. Natl. Acad. Sci. USA.* 85:8156.
14. Clevers, H., N. Lonberg, S. Dunlap, E. Lacy, and C. Terhorst. 1989. An enhancer located in a CpG-island 3′ to the TCR/CD3-ϵ gene confers T lymphocyte-specificity to its promoter. *EMBO (Eur. Mol. Biol. Organ.) J.* 8:2527.
15. Clevers, H., B. Alarcon, T. Wileman, and C. Terhorst. 1988. The TCR/CD3 complex: a dynamic protein ensemble. *Annu. Rev. Immunol.* 6:629.
16. van de Wetering, M., M. Oosterwegel, D. Dooijes, and H. Clevers. 1991. Identification and cloning of TCF-1, a T cell-specific transcription factor containing a sequence-specific HMG box. *EMBO (Eur. Mol. Biol. Organ.) J.* In press.
17. Jantzen, H.-M., A. Admon, S.P. Bell, and R. Tjian. 1990. Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins. *Nature (Lond.).* 344:830.
18. Sinclair, A.H., P. Berta, M.S. Palmer, J.R. Hawkins, B.L. Griffiths, M.J. Smith, J.W. Foster, A.-M. Frischau, 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 (Lond.).* 346:240.
19. Kelly, M., J. Burke, M. Smith, A. Klar, and D. Beach. 1988. Four mating-type genes control sexual differentiation in the fission yeast. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:1537.
20. Staben, C., and C. Yanofsky. 1990. Neurospora crassa a mating-type region. *Proc. Natl. Acad. Sci. USA.* 87:4917.
21. Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular Cloning: A Laboratory Manual. 2nd edition. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
22. Wilkinson, D.G., J.A. Bailes, J.E. Champion, and A.P. McMahon. 1987. A molecular analysis of mouse development from 8 to 10 days post coitum detects changes only in embryonic globin expression. *Development.* 99:493.
23. Kress, C., R. Vogels, W. de Graaff, C. Bonnerot, F. Meijlink, J.-F. Nicolas, and J. Deschamps. 1990. Hox-2.3 upstream sequences mediate lacZ expression in intermediate mesoderm derivatives of transgenic mice. *Development (Camb.).* 109:493.
24. Uhlen, M., B. Nilsson, B. Guss, M. Lindberg, S. Gatenbeck, and L. Philipson. 1983. Gene fusion vectors based on the gene for staphylococcal protein A. *Gene (Amst.).* 23:369.
25. Georgopoulos, K., D. Galson, and C. Terhorst. 1990. Tissue-specific nuclear factors mediate expression of the CD3-δ gene during T cell development. *EMBO (Eur. Mol. Biol. Organ.) J.* 9:109.
26. Winoto, A., and D. Balimore. 1989. A novel, inducible and T cell-specific enhancer located at the 3′ end of the T cell receptor α locus. *EMBO (Eur. Mol. Biol. Organ.) J.* 8:729.
27. Ho, I.-C., L.-H. Yang, G. Morle, and J.M. Leiden. 1989. A T cell-specific transcriptional enhancer element 3′ of Cα in the human T-cell receptor α locus. *Proc. Natl. Acad. Sci. USA.* 86:6714.
28. Ho, I.-C., and J.M. Leiden. 1990. 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.
29. Waterman, M.L., and K.A. Jones. 1990. Purification of TCF-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.
30. Mermod, N., E.A. O'Neill, T.J. Kelly, and R. Tjian. 1989. The proline-rich transcriptional activator of CTF/NF-1 is distinct from the replication and DNA binding domain. *Cell.* 58:741.
31. 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.