The EMBO Journal vol.10 no.1 pp.123–132, 1991

Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box

Marc van de Wetering, Mariette Oosterwegel, Dennis Dooijes and Hans Clevers

Department of Clinical Immunology, University Hospital Utrecht, PO Box 85500, 3508 GA, Utrecht, The Netherlands

Communicated by P.Borst

***CD3-ϵ* gene expression is controlled by a downstream T lymphocyte-specific enhancer element. We report the identification of a T cell-specific transcription factor, TCF-1, binding to this element. The multimerized recognition motif of TCF-1 constituted a T cell-specific enhancer. Subsequent cloning of TCF-1 identified three splice alternatives. TCF-1 contained a single DNA-binding HMG box most closely related to similar boxes in the putative mammalian sex-determining gene SRY and in the *Schizosaccharomyces pombe* Mc mating type gene. TCF-1 mRNA was expressed uniquely in T lymphocytes. Upon cotransfection into non-T cells, TCF-1 could transactivate through its cognate motif. These results identify TCF-1 as a T cell-specific transcription factor, which might play a role in the establishment of the mature T cell phenotype.**

Key words: CD3 genes/HMG box/T cell-specific transcription factor

Introduction

In recent years, much information has been obtained on genes that control the phenotype of cells within metazoan organisms. The genetic identification of genes involved in pattern formation of *Drosophila melanogaster* and *Caenorhabditis elegans* has been particularly informative. Many of these genes appear to encode transcription factors. This has led to the notion that the control of cellular phenotype principally occurs at the level of transcription. Developmentally active transcription factors regulate the expression of sets of structural genes, and are thus responsible for the induction and maintenance of a particular phenotype (Maniatis *et al.*, 1987; Scott and Carroll, 1987; Atchison, 1988; Johnson and McKnight, 1989).

Obvious practical limitations complicate similar developmental studies of higher eukaryotes, especially of mammals. Such studies are therefore often conducted from a different perspective. Rather than analysing early embryonic events involved in pattern-formation, attention has focused on the terminal differentiation of precursor cells into their functionally mature counterparts. It has become feasible to clone genes that are uniquely expressed in those end-stage cells, and to identify the *cis*-acting elements that target transcription of such genes to particular differentiation stages. These *cis*-acting elements (promoters, enhancers, silencers) can subsequently be used as tools to identify transcription factors that bind to specific sequence motifs within these elements. Of particular interest are transcription factors that are uniquely active in the cell type where the structural, end-stage gene is expressed. A number of tissue-specific transcription factors have thus been identified and cloned, including the pituitary factor Pit-1, which controls growth hormone and prolactin gene expression (Bodner *et al.*, 1988; Ingraham *et al.*, 1988); the B cell factor Oct-2, involved in expression of immunoglobulin genes (Mueller *et al.*, 1988; Scheidereit *et al.*, 1988); the erythrocyte factor EryFI, binding to haemoglobin regulatory sequences (Evans and Felsenfeld, 1989; Tsai *et al.*, 1989); and the hepatic transcription factors LF-BI (Frain *et al.*, 1989) and DBP (Mueller *et al.*, 1990). Probably the best documented example of a transcription factor controlling cell fate in higher eukaryotes is that of *MyoD* and related genes. *MyoD* has been shown to confer the muscle cell phenotype onto fibroblasts (Davis *et al.*, 1987) and to activate muscle-specific genes in cells of various origins (Weintraub *et al.*, 1989). Subsequently, *MyoD* was demonstrated to encode a DNA-binding protein, capable of binding and transactivating the muscle-specific muscle creatine kinase enhancer (Lassar *et al.*, 1989).

Little is known about the molecular events that govern the differentiation pathway of the T lymphocyte. One of the early events in T lineage commitment is the expression of the members of the CD3 complex. This complex consists of at least five invariable, integral membrane proteins and is non-covalently associated with the T cell antigen receptor (TCR) on the surface of mature T cells. Three of the *CD3* genes (*CD3-γ*, *CD3-δ*, *CD3-ϵ*) form a small family. They are located on a stretch of 60 kb on human chromosome 11q23. Their expression is uniquely restricted to cells of the T lineage and precedes rearrangement and expression of the *TCR* genes (Clevers *et al.*, 1988a; Tunnacliffe *et al.*, 1988). The tight clustering of the *CD3-γ/δ/ϵ* locus might suggest that the *CD3* genes are controlled by a single *cis*-acting element. This, however, appears not to be the case. Transgenic mice, carrying either the human *CD3-δ* or the human *CD3-ϵ* gene on non-overlapping fragments, expressed the transgene correctly, demonstrating that the two genes each carry a complete and independent set of regulatory elements (Clevers *et al.*, 1989; H.Clevers and N.Lonberg, unpublished).

In order to identify factors that specify the T lymphocyte phenotype, we have set out to characterize T cell-specific DNA-binding proteins involved in the tissue-specific expression of the *CD3-ϵ* gene. The human *CD3-ϵ* gene consists of nine exons and spans 12 kb (Clevers *et al.*, 1988b). T cell-specific expression of *CD3-ϵ* results from an enhancer located directly downstream of exon IX. This enhancer occupies ~1.5 kb, constitutes a CpG island, and coincides with a DNase I-hypersensitive site present only in T lymphocytes. Within the *CD3-ϵ* enhancer, a 130 bp core is present, which by itself functions as a T cell-specific enhancer; deletion of this core renders the remaining

enhancer sequences inactive (Clevers *et al.*, 1989). Here we report the identification and cloning of a transcription factor, TCF-1, which is present in T cells only and which recognizes a motif in the *CD3-ϵ* enhancer.

Results

Identification of a T cell-specific DNA-binding protein, TCF-1

To identify putative transcription factors responsible for the T cell-specificity of the *CD3-ϵ* enhancer, crude nuclear extracts from a panel of human T and non-T cell lines were analysed for the presence of specific DNA-binding proteins by gel retardation. A set of four overlapping probes was generated from the *CD3-ϵ* enhancer utilizing two internal restriction sites (Figure 1A and 1B). A number of sequence-specific complexes could be identified with these probes. One of these complexes was present in all T cell lines and absent from all other cells tested. This T cell-specific complex was only observed with probe L, giving an indication of the actual site of contact. Figure 1C depicts a representative experiment with the probes S and L using a nuclear extract from the human T cell line Jurkat. The T cell-specific DNA−protein complex appears as a doublet.

In order to map the recognition motif more closely, three overlapping double-stranded oligonucleotides were synthesized (MW12: bp 99−123; MW34: bp 86−110; and MW56: bp 70−101). Gel retardation analysis performed with these three oligonucleotide probes revealed that the T cell-specific protein−DNA complex was formed only with MW56, and again appeared as a typical doublet. Figure 2A depicts an overexposure of a retardation experiment performed with probe MW56 on a large panel of nuclear extracts. As is evident, a prominent retarded band was observed exclusively with nuclear extracts of all T cell lines. A similar retarded band was found with extracts of the murine T cell line EL4, but not with the murine B lineage cell line Ag8, nor with the murine fibroblast cell line Swiss-3T3 (data not shown). The T cell-specific binding activity was termed TCF-1.

The binding specificity of TCF-1 for MW56 was confirmed in competition experiments. Excess unlabelled MW56 competed efficiently for the doublet band formed with the MW56 probe, whereas excess MW12 and MW34 had no effect (Figure 3). Furthermore, excess cold MW56 (but not MW12 or MW34) could block formation of the doublet formed with probe L, proving that the T cell-specific doublets observed with probe L and with MW56 resulted from the same binding activity (not shown).

To define the nucleotides contacted by TCF-1, methylation interference footprinting was performed with T cell nuclear extracts using MW56, end-labelled at either the positive (MW5) or negative (MW6) strand. Experimental conditions were such that interactions with A and G appeared as spots of decreased intensity; interactions with C and T were not analysed. As depicted in Figure 4, the positive strand footprint of Jurkat TCF-1 consisted of 5′gggagactg-agAAcAAAGcgctctcacac (where contact bases are indicated in capitals). No contacts were observed with G and A on the negative strand. Footprinting experiments performed with nuclear extracts from the human T cell line H9 confirmed these results (data not shown).

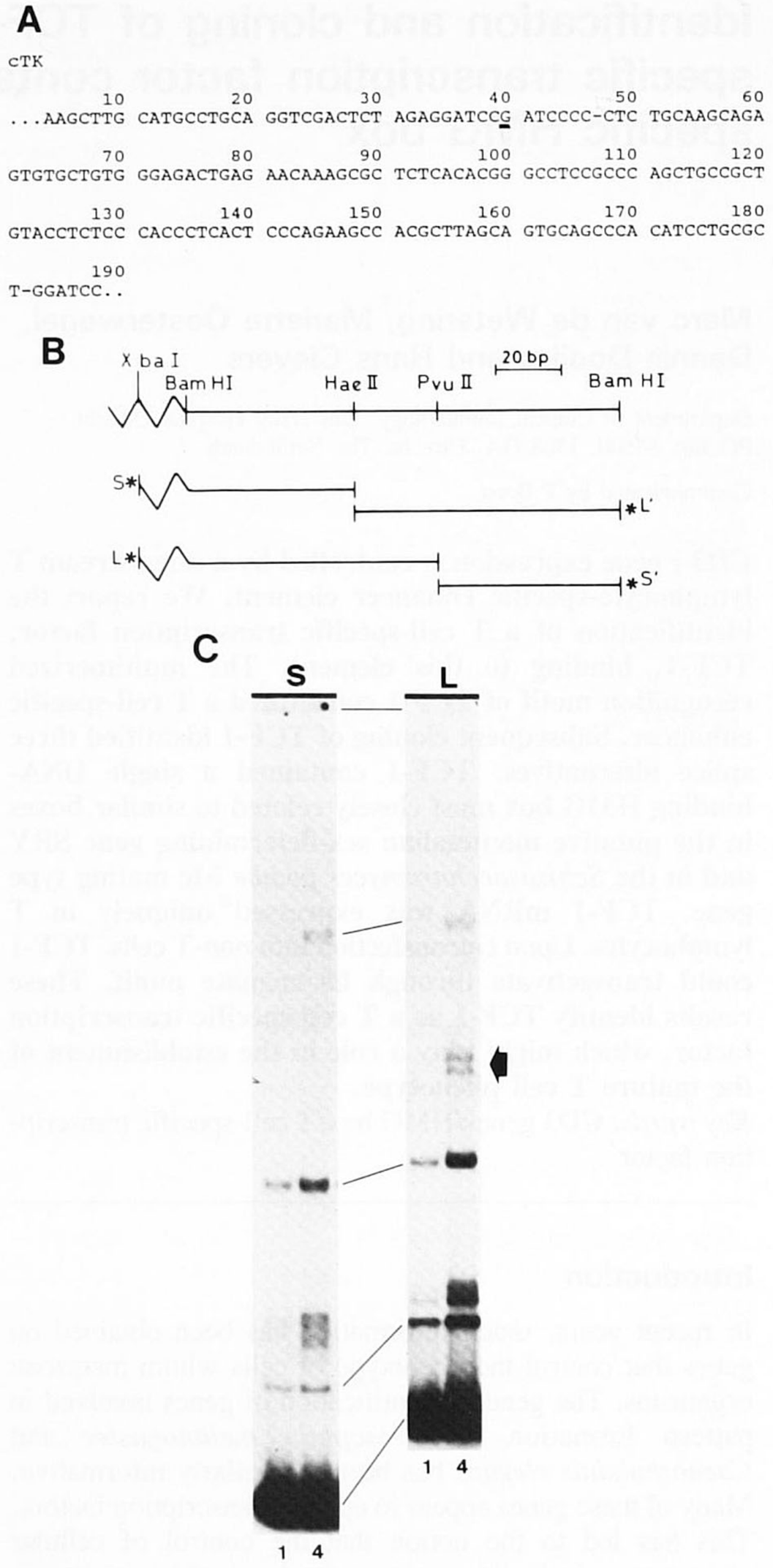


Fig. 1. Gel retardation analysis performed on the *CD3-ϵ* enhancer. **A.** The T lymphocyte-specific core of the *CD3-ϵ* enhancer, as present in deletion clone pI9-c (Clevers *et al.*, 1989) was subcloned in the *Bam*HI site of the pBLCAT2 polylinker (see Materials and methods section), yielding pcTK. The two hyphens surround the actual enhancer sequence. Additional bases result from subcloning procedures. **B.** Four probes were generated from pcTK using the indicated restriction sites. *Xba*I at bp 28; *Eco*47III at bp 86; *Bam*HI at bp 34 and 183; and *Pvu*II at bp 110. **C.** Gel retardation analysis performed with the probes S (left) and L (right) using 1 or 4 μg of Jurkat nuclear protein. The arrow indicates the T cell-specific DNA−protein complex TCF-1 observed with probe L.

The multimerized TCF-1 motif constitutes a T cell-specific enhancer

It was demonstrated next that the TCF-1 motif, taken out of the context of the *CD3-ϵ* enhancer, could still activate transcription in a T cell-specific fashion. To that end, a

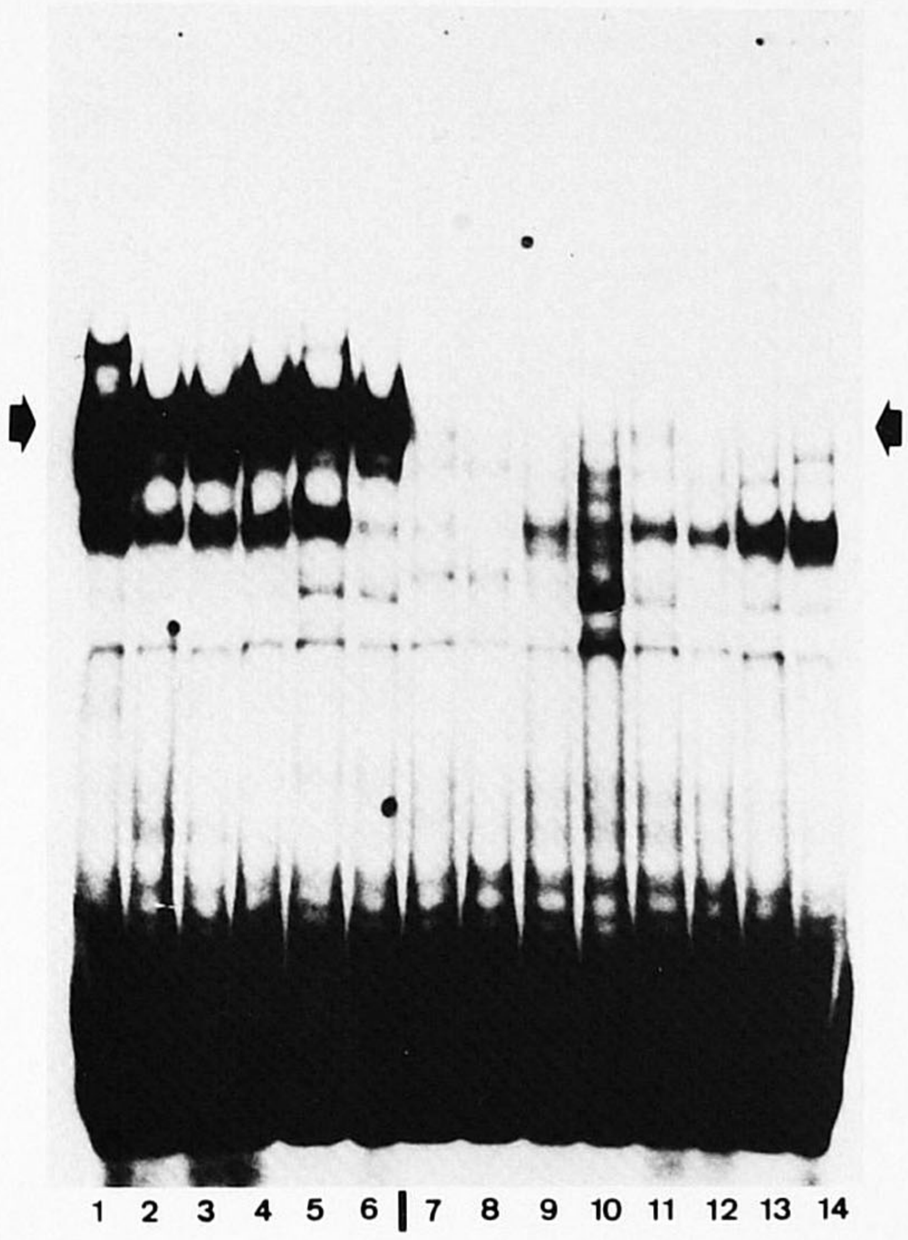


Fig. 2. Cell-type specificity of TCF-1. Overexposure of a gel retardation analysis performed on a panel of human T cell lines (lanes 1–6) and non-T cell lines (lanes 7–14) with the double-stranded oligonucleotide MW56 (bp 69–101 of pcTK) as probe. 1: CCRF-CEM; 2: CCRF-HSB; 3: H9; 4: Jurkat; 5: Molt-4; 6: Peer; 7: CRL-1484; 8: Daudi; 9: N-67; 10: BS-5.2; 11: K-562; 12: Meg-01; 13: HeLa; 14: *Hep*G2. The arrows indicate the position of TCF-1.

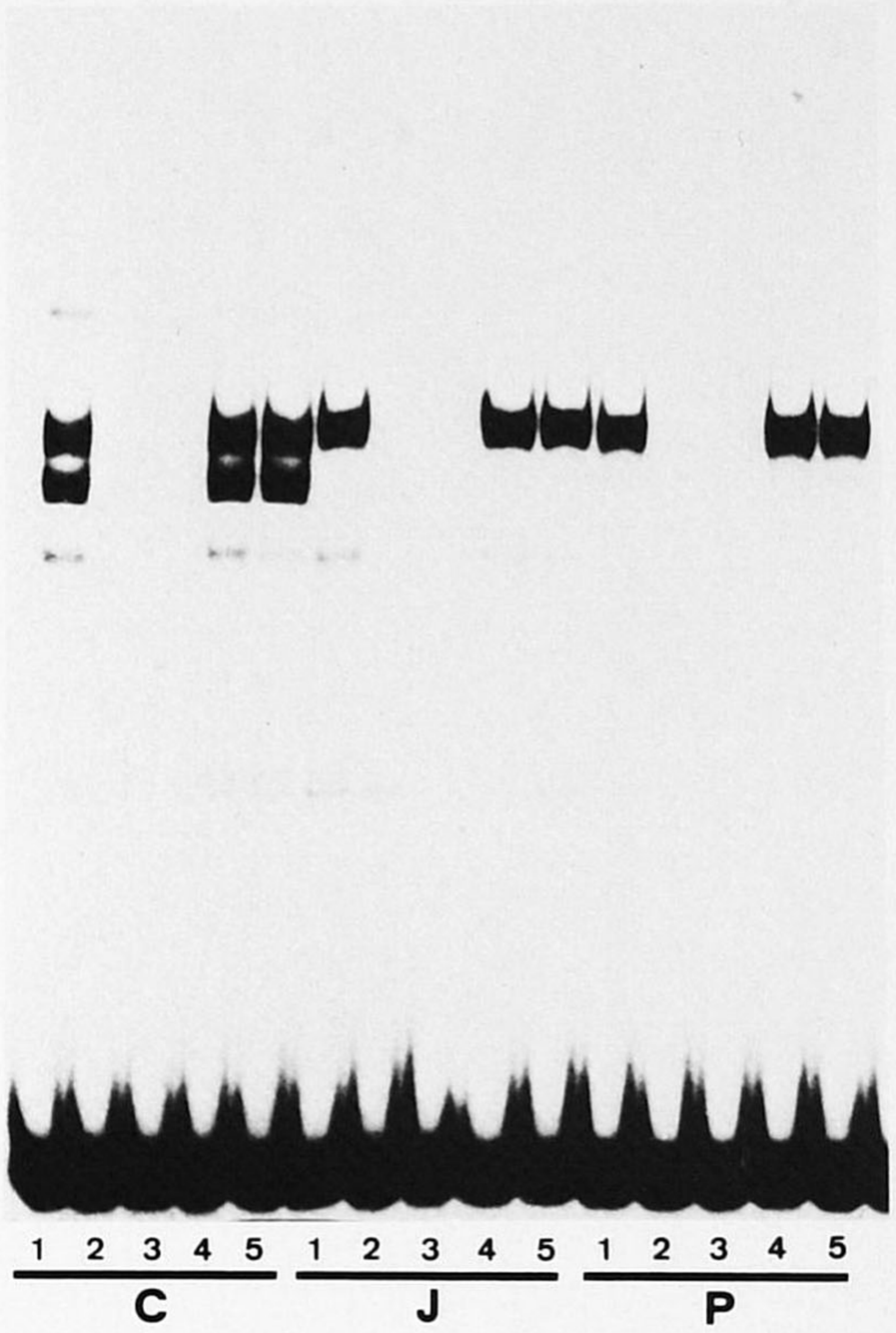


Fig. 3. Sequence-specificity of TCF-1. Gel retardation analysis was performed with extracts of the T cell lines CCRF-CEM (C), Jurkat (J), and Peer (P). Lanes 1: no competitor DNA; lanes 2: 50 ng of cold MW56; lanes 3: 100 ng of cold MW56; lanes 4: 100 ng of cold (irrelevant) MW12; lanes 5: 100 ng of cold (irrelevant) MW34. Remarkably, in all extracts prepared from the CCRF-CEM, a second doublet was observed running slightly below the TCF-1 doublet (Figures 2 and 3). The nature of this second doublet is so far unknown.

concatamer of seven copies of MW56 was cloned upstream of the minimal herpes simplex virus thymidine kinase (TK) promoter of pBLCAT2 (Luckow and Schutz, 1987); the resulting construct was termed pMW$(56)_7$. The presence of the MW56-concatamer led to a 5-fold increase of the basal TK activity upon transfection in Jurkat T cells as measured in a CAT assay (see Table I). CAT analysis performed with non-T cell lines proved this enhancement to be T cell-specific (Table I).

To confirm that the TCF-1 recognition motif was responsible for the observed enhancement of TK promoter activity, a variant of MW56 was synthesized in which the AACAAAG footprint was replaced by CCGCGGT (MW56Sac). As expected, a concatamer containing seven copies of MW56Sac (pMW56Sac$_7$) failed to increase the activity of the TK promoter in T cells (Table I). Control experiments showed that MW56Sac did not bind TCF-1 in the gel retardation assay, nor did cold excess MW56Sac compete with MW56 for the binding of TCF-1 (data not shown).

These experiments lent strong support to the notion that the interaction of TCF-1 with its cognate sequence activates a linked promoter, thus identifying TCF-1 as a transcription factor.

Molecular cloning of TCF-1

To be able to study TCF-1 more precisely and gain insight into its involvement in T lymphocyte development, cDNA clones encoding TCF-1 were isolated. A λgt11 cDNA library derived from the human T cell line Jurkat was screened with multimerized double-stranded MW56 as a probe (Singh *et al.*, 1988; Vinson *et al.*, 1988). A primary screen of 2×10^6 recombinant phages yielded one positive clone, designated ΦTCF-1. Initial characterization of this recombinant phage revealed that the β-galactosidase fusion protein specifically bound the MW56 probe, and failed to bind multimerized MW12 and MW34 probes. Furthermore, binding of the MW56 probe to the fusion protein was blocked by excess cold MW56, but not by MW12, MW34, or calf thymus DNA (data not shown).The 0.9 kb insert of this phage clone was then used to screen the same Jurkat library, and a plasmid cDNA library derived from the human T cell line HPB-ALL. Three additional phage clones and 31 recombinant plasmids were isolated. The isolated cDNA clones could be divided into three groups as determined by restriction pattern analysis.

Sequence analysis of the ΦTCF-1 insert and of representative cDNA clones isolated in the second screen yielded the following information. Two in-frame methionines were preceded by a stop codon at bp 9 of the longest clone. The translation initiation site was tentatively placed at the first ATG codon (bp 80), showing the best match with the Kozak-consensus (Kozak, 1984). The N-terminal part of the

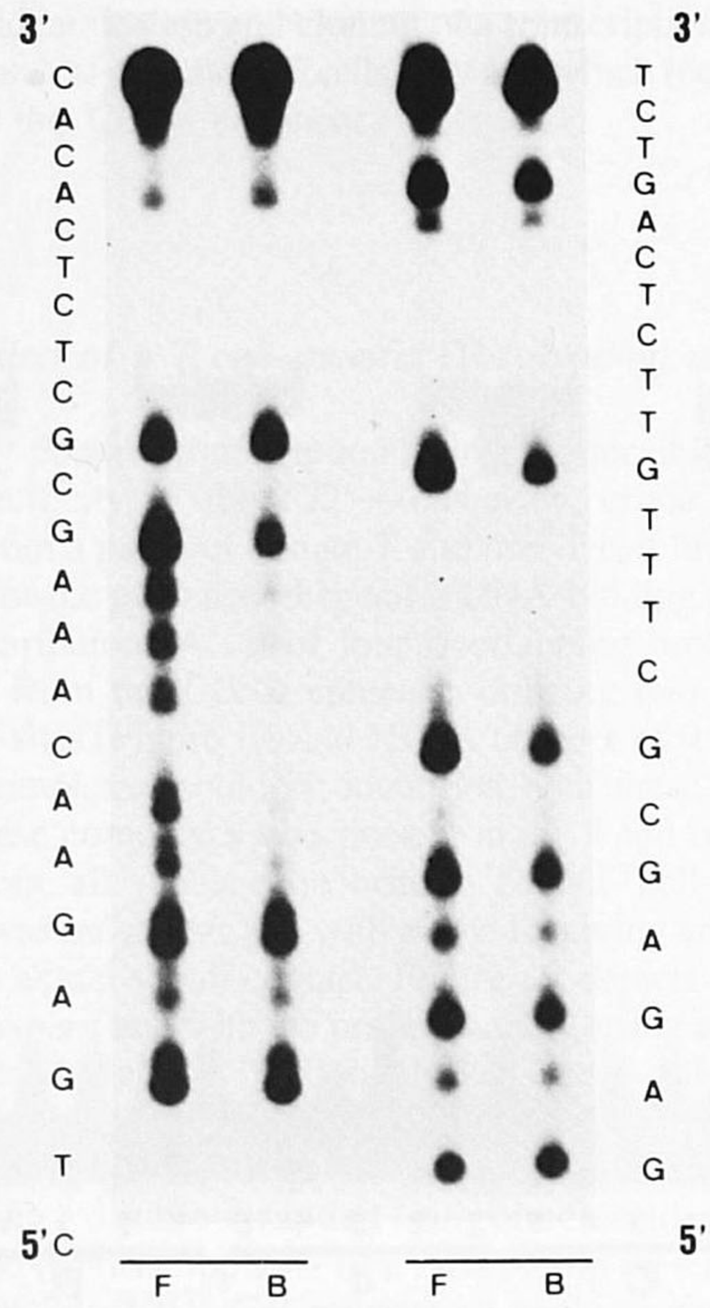


Fig. 4. Methylation interference footprinting of TCF-1 on the MW56 probe. The left panel: coding strand; right panel: non-coding strand. F: reaction performed on free probe eluted after gel retardation analysis; B: reaction performed on TCF-1-bound probe eluted after gel retardation analysis. A clear footprint was present on the coding strand only: 5′ AA(c)AAAG 3′

predicted amino acid sequence (residues 17−124) was relatively rich in proline residues (22%). This characteristic is not uncommon in transcription factors, and has been demonstrated in one case at least to constitute a transcription activation domain (Mermod *et al.*, 1989; and references therein). The second part of the sequence shared by all clones (residues 125−243) was highly charged, and predominantly basic, a feature indicative of a DNA binding domain. Indeed, this domain was contained within the insert of ΦTCF-1 (Figure 5).

The sequence of the different types of cDNA clones diverged downstream from bp 838. The presence of the consensus bases AG at the end of the shared sequence fits well with the occurrence of alternative splicing at this position. The three alternative splice forms were termed TCF-1A, TCF-1B and TCF-1C. The TCF-1A splice form, represented by the insert of ΦTCF-1, was not isolated again in the second screen. The alternative exon of TCF-1A predicted the addition of 25 amino acid residues at the C-terminus of TCF-1, resulting in a total length of 269 amino acids. The alternative exon of TCF-1B similarly predicted the addition of 25 C-terminal amino acids; however, no homology to TCF-1A was apparent in this area. In clones of type TCF-1C, a stretch of 96 bp (ending with AG), was inserted 5′ of the TCF-1B alternative exon. This insertional exon added 24 amino acids to the TCF-1 sequence, and precluded translation of the C-terminus encoded by TCF-1B. The TCF-1C encoded protein was predicted to consist of

Table I. The TCF-1 cognate motif constitutes a T cell-specific transcription element

	Jurkat	Daudi	K562	HeLa	HepG2
pBLCAT2	3063	264	1715	856	7003
pTKRSV	17247	1760	33118	38010	21755
pcTK	14299	229	2285	432	5998
$pTK56_7$	14613	230	2392	222	6256
$pTK56Sac_7$	1601	–	–	–	–

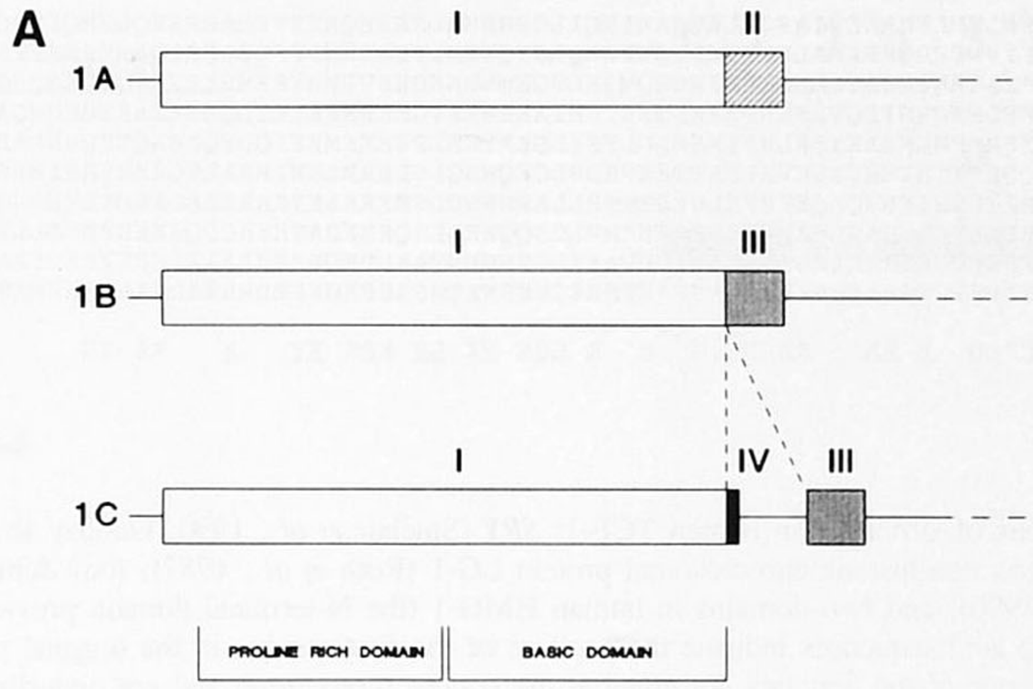
The indicated cells were transfected with 10 μg of CAT plasmid and cell lysates were tested for CAT activity 48 h later. The results are represented as c.p.m. of extractable (butyrylated) [^{14}C]chloramphenicol. pBLCAT2 contains the CAT gene driven by the non-tissue specific TK promoter (Luckow and Schutz, 1987). The presence of the RSV enhancer clearly augments basal activity of the TK promoter in all cell types (pTKRSV, Clevers *et al.*, 1989). The *CD3-ϵ* enhancer is active in the T cell line only (pcTK in Jurkat). Similarly, seven copies of an oligonucleotide containing the TCF-1 cognate motif enhance the activity of the TK promoter in Jurkat cells only ($pTK56_7$). Replacement of the TCF-1 contact bases in this oligonucleotide by irrelevant sequence abrogates this T cell-specific enhancer activity ($pTK56Sac_7$).

268 amino acid residues. Preliminary characterization of genomic TCF-1 clones has confirmed the exon/intron boundary at bp 838, and has demonstrated that all alternative exons map within an area of 5 kb (M. van de Wetering, F.Holstege, and H.Clevers, unpublished).

Since no additional cDNA clones of the TCF-1A type were isolated, it was expected that, at least in Jurkat and HPB-ALL, TCF-1A represents a relatively rare splicing event. Based on the longest TCF-1B and TCF-1C clones, the sizes of the corresponding mRNAs were predicted to be at least 2.9−3.0 kb, which fitted well with the 3.0 kb size estimated from Northern blotting (see below). In two clones, an imperfect alternative polyadenylation signal (AATTAA at bp 1348 of TCF-1B) appeared to be utilized, which resulted in the addition of a poly-A tail at bp 1365 and at bp 1368 respectively. Indeed, overexposures of Northern blot experiments revealed the presence of mRNA species of ~1400−1500 bp (not shown). An Alu repeat was present between bp 2291 and 2576 of the 3′ untranslated sequence of TCF-1B. A consensus polyadenylation signal (AATAAA at bp 2796) was found near the 3′ end of the longest cDNA clone, and was followed by a stretch of 7 As at bp 2836 of the longest clone.

TCF-1 is a member of an 'HMG box-family' of DNA-binding proteins

The primary TCF-1 sequence exhibited none of the amino acid motifs (leucine-zipper, helix-loop-helix, zinc-finger, homeobox) that have been observed in many recently identified eukaryotic transcription factors (Johnson and McKnight, 1989). Computer-assisted searches of NBRF and Swiss Protein Sequence Databases using the Lipman/Pearson algorithm (Lipman and Pearson, 1985) revealed a significant similarity of TCF-1 with Mc, one of the mating type genes of *Schizosaccharomyces pombe*, the fission yeast (23% similarity over a stretch of 77 amino acids). A slightly lower degree of similarity (20% over the same stretch) was obtained for a match with *Tetrahymena* LG-1, a high mobility group-1 (HMG-1)-related protein. Similarities of borderline significance were found with *HMG*-1 genes of various species. Database searches with Mc and LG-1 detected reciprocal similarities. Alignment of TCF-1, Mc,



B

```
I  gcccaggtgactgactaatccgccgccttcaggagacagaattggccaaggcctgaag   58

gccccggagtgcaccagcggcATGTACAAAGAGACCGTCTACTCCGCCTTCAATCTGCTC 118
                    MetTyrLysGluThrValTyrSerAlaPheAsnLeuLeu  13

ATGCATTACCCACCCCCCTCGGGAGCAGGGCAGCACCCCCAGCCGCAGCCCCCGCTGCAC 178
MetHisTyrProProProSerGlyAlaGlyGlnHisProGlnProGlnProProLeuHis  33

AAGGCCAATCAGCCCCCCCACGGTGTCCCCCAACTCTCTCTCTACGAACATTTCAACAGC 238
LysAlaAsnGlnProProHisGlyValProGlnLeuSerLeuTyrGluHisPheAsnSer  53

CCACATCCCACCCCTGCACCTGCGGACATCAGCCAGAAGCAAGTTCACAGGCCTCTGCAG 298
ProHisProThrProAlaProAlaAspIleSerGlnLysGlnValHisArgProLeuGln  73

ACCCCTGACCTCTCTGGCTTCTACTCCCTGACCTCAGGCAGCATGGGGCAGCTCCCCCAC 358
ThrProAspLeuSerGlyPheTyrSerLeuThrSerGlySerMetGlyGlnLeuProHis  93

ACTGTGAGCTGGTTCACCCACCCATCCTTGATGCTAGGTTCTGGTGTACCTGGTCACCCA 418
ThrValSerTrpPheThrHisProSerLeuMetLeuGlySerGlyValProGlyHisPro 113

GCAGCCATCCCCCACCCGGCCATTGTGCCCCCCTCAGGGAAGCAGGAGCTGCAGCCCTTC 478
AlaAlaIleProHisProAlaIleValProProSerGlyLysGlnGluLeuGlnProPhe 133

GACCGCAACCTGAAGACACAAGCAGAGTCCAAGGCAGAGAAGGAGGCCAAGAAGCCAACC 538
AspArgAsnLeuLysThrGlnAlaGluSerLysAlaGluLysGluAlaLysLysProThr 153

ATCAAGAAGCCCCTCAATGCCTTCATGCTGTACATGAAGGAGATGAGAGCCAAGGTCATT 598
IleLysLysProLeuAsnAlaPheMetLeuTyrMetLysGluMetArgAlaLysValIle 173

GCAGAGTGCACACTTAAGGAGAGCGCTGCCATCAACCAGATCCTGGGCCGCAGGTGGCAC 658
AlaGluCysThrLeuLysGluSerAlaAlaIleAsnGlnIleLeuGlyArgArgTrpHis 193

GCGCTGTCGCGAGAAGAGCAGGCCAAGTACTATGAGCTGGCCCGCAAGGAGAGGCAGCTG 718
AlaLeuSerArgGluGluGlnAlaLysTyrTyrGluLeuAlaArgLysGluArgGlnLeu 213

CACATGCAGCTATACCCAGGCTGGTCAGCGCGGGACAACTACGGGAAGAAGAAGAGGCGG 778
HisMetGlnLeuTyrProGlyTrpSerAlaArgAspAsnTyrGlyLysLysLysArgArg 233

TCGAGGGAAAAGCACCAAGAATCCACCACAG                              838
SerArgGluLysHisGlnGluSerThrThr                               243
```

C

```
II AGACAAACTGGCCCAGAGAACTCAAGGATGGTAATGGACAAGAGTCACTGTCCATGTCT  897
   GluThrAsnTrpProArgGluLeuLysAspGlyAsnGlyGlnGluSerLeuSerMetSer  263

   TCTTCCTCTAGCCCAGCTtgaggactgggatggctgggcaaggaagccataggcattgcg  957
   SerSerSerSerProAla                                            269

   gccccttgccttggtgcagatgtgagtcccacaaacacatctggagaagctcaaaggccg 1017
   ggactgggagatgactcccttggaagacaggagagatgactcccttggaagacagatgac 1077
   agcccataggcctagtgacaaaaggccccctttccgaccttgtggctgttctgggaactgc 1133
   acctgtcctaggtctgggccagaccaagcagaatggcagtctgaggacactgacttacca 1193
   cccaagtcccaggaagagaggacaaggaatcagccaggcctgtgcaaaggcagcattttt 1253
   tggttgtggagtatgactatgaattc                                   1279

III GAGGAAAAAGAAATGCATTCGGTACTTACCCGGAGAAGGCCGCTGCCCCAGCCCCGTTC 897
   GlyGlyLysArgAsnAlaPheGlyThrTyrProGluLysAlaAlaAlaProAlaProPhe  263

   CTTCCGATGACAGTGCTCtaggctgccccgggtccccagctccccaggactcaccctcat  957
   LeuProMetThrValLeu                                            269

   accatctgctgccccgcttccccacagaactgcttactagccctgcggagcggcacctac 1017
   atccccaggtctctccactgctctcagcctcccaaccccagggcccccacaggcccccg  1077
   cagcaccctgcagagcacacaggtacagcaacaggaatctcagagacaggtggcctagca 1137
   ggcacaggacacctggccgcctccaggagcctacccccctgaaaatgacagagacccagat 1197
   ctcatggaaactggcaggggtcctgttaacgtcatctcagggtccagaccctgaagattt 1257
   cagaggctgcagaacttctgcctgaacctggggtcatcgattcaaactgctccaagtggt 1317
   gggaatcagatctgtcttgatgtgtcatctaattaagggaatcccttgtacctatgactc 1377
   cgtcgatctattctttgtaccatctgtcttgccagccagaagcctctgcctccctagctt 1437
   ttctgctataggtcagagatgggctgaactgagcctagctaccttctctacccatctccc 1497
   ccatcccccactgccacaccctccccattcagacacttcatggaccaagaatgagctggt 1557
   ttgtcaaacaacatgtgagcatggtcacaagcacaaagctcaagatgacagctcttctaa 1617
   ggaaatggagaagctctgtttataaaaacaaaaacaaaaccagctgctactcataagttg 1677
   gaccagaggaagccccttactatgatctcaggagcttgcaagaagcaggaaggggaatgg 1737
   aataggttaagtttaggcctatcaacctaagcaacagaaataatctgacactaccttatc 1797
   aggcaaattggggagggagggtgtatctagctctagttcaaattatttggaagtgttcc 1857
   ctgagaaacccaccagcctaagaagctctggccccaggcttgtcactagcagctgcagtc 1917
   aacagttcaaagaagtcatggcccaaatccagtgtgcaccctccccattcacagagcct 1977
   ttttcacaattccatttccagttcatctatggcagtccagccagcttctgggcagcttga 2037
   gagggcaaacccaaaacctcatgacagccagagcctgtctttcagcattcagtcgcctgg 2097
   gaagctccagtttccccatggggctgcgggacagaggaccattacaactagatcaaggag 2157
   cccagaaaacctccagtagtggacaacaggttttcaccatagcctacgttaacccatttt 2217
   tgagccaagcttcaaccctcagccttgaaaaacaagtctttaatttaatttttgttttttt 2277
   gcctaaatccaaagaaaaaggctgccgggccaggcgcggtcgtcacgcctgtaatccca 2337
   gcactttggcaggccgaggcaggtggatcacctgacgtcagtagtttgagaccagcctgg 2397
   ccaacatggtgaaaccctgtatctactaaaaatacaaaaattagccggacgtggtggtgc 2457
   gcgcatgtaatcccagctactcgggaggctgaggcggaagaatcacttgaacccgggagg 2517
   cggaggttccagtgagccgagatggcgctattgcactccagtctgggtaacagggagact 2577
   gcatctcaaaaaaaaaaaaaaaaaaaaaggctgtccatgtattcacacacccctcaaaa 2637
   aaagcctttagttcctactttagccactggtttctcagaatccaaagatcacatattcta 2697
   gtgtaacactgcaagaagtcttgagaaaaagattattgtagtgttcaaaatatttttgta 2757
   ttgttaatgcatcatcatagaaaaactttaaacatgagaataaagatacttttttactgg 2817
   gtttgtttttcaaagcctgaaaaaaa                                   2843

IV ACCCTGGCTCGCCTAAGAAATGCCGTGCTCGCTTTGGCCTCAACCAGCAGACGGATTGG  897
   AspProGlySerProLysLysCysArgAlaArgPheGlyLeuAsnGlnGlnThrAspTrp  248

   TGTGGTCCGTGCAGAtaactctcttcactattcctag                        934
   CysGlyProCysArg                                              268
```

Fig. 5. Sequence of three alternative splice forms of TCF-1. **A.** All three splice forms share the proline rich and the DNA-binding basic region (I). The clones differ at their C-terminus. TCF-1A utilizes sequence II; TCF-1B utilizes sequence III; and TCF-1C contains the 97 bp sequence IV inserted between sequence I and III. No full length TCF-1A clone was isolated in this study. The 5′ boundary of the partial TCF-1A clone is indicated in 5B. Upstream sequences were extrapolated from full-length TCF-1B and -1C clones. **B.** Nucleotide and deduced amino acid sequences of the shared TCF-1 region I. An asterisk marks an in-frame stop codon, preceding the long open reading frame. The arrow indicates the start of the original phage clone ΦTCF-1. **C.** Alternative nucleotide and deduced amino acid sequences of regions II, III and IV. Polyadenylation signals in III are underlined.

LG-1, and human HMG-1 revealed that the region of similarity was shared between all four proteins. Of interest, this region mapped within the TCF-1 DNA-binding domain as defined by the insert of ΦTCF-1 (amino acids 102−243).

Three recent studies extend the number of genes with similarity to this HMG-1 domain. Tjian and colleagues report the cloning of the polymerase I transcription factor UBF, and demonstrate the presence of four repeats of a sequence-specific DNA-binding domain within UBF. This repeated domain was homologous to a region in *HMG*-1, and was therefore termed the HMG box (Jantzen *et al.*, 1990). In two other reports, a candidate sex-determining gene, *SRY*,

```
TCF-1: 154 IKKPLNAFMLYMKEMRAKVIAECTLKESAAINQILGRRWHALSREEQAKYYELARKERQLHMQLYPGW--SA-RDNYGKK
SRY  :     VKRPMNAFIVWSRDQRRKMALENPRMRNSEISKQLGYQWKMLTEAEKWPFFQEAQKLQAMHREKYPNY--KY-RPRRKAK
Mc   : 103 TPRPPNAFILYRKEKHATLLKSNPSINNSQVSKLVGEMWRNESKEVRMRYFKMSEFYKAQHQKMYPGY--KY-QPRKNKV
LG-1 : 123 PKRPLSAFFLFKQHNYEQVKKENPNAKITELTSMIAEKWKAVGEKEKKKYETLQSEAKAKYEKDMQAYEKKYGKPEKQKK
UBF 1: 112 PKKPLTPYFRFFMEKRAKYAKLHPEMSNLDLTKILSKKYKELPEKKKMKYIQDFQREKQEFERNLARFREDHPDLIQNAK
UBF 2: 196 PEKPKTPQQLWYTHEKKVYLKVRPDATTKEVKDSLGKQWSQLSDKKRLKWIHKALEQRKEYEEIMRDYIQKHPELNIS
UBF 3: 406 PKRPVSAMFIFSEEKRRQLQEERPELSESELTRLLARMWNDLSEKKKAKYKAREAALKAQSERKPGGEREERGKLPESPK
UBF 4: 294 DGRPTKPPPNSYSLYCAELMANMKDVPSTERMVLCSQQWKLLSQKEKDAYHKKCDQKKKDYEVELLRFLESLPEEEQQRV
HMG-1:  94 PKRPPSAFFLFCSEYRPKIKGEHPGLSIGDVAKKLGEMWNNTAADDKQPYEKKAAKLKEKYEKDIAAYRAK-GKPDAAKK
HMG-1:   8 PRGKMSSYAFFVQTCREEHKKKHPDASV*EFSKKCSERWKTMSAKEKGKFEDMAKADKARYEREMKTYIPPKGETKKKFK

consensus PKKP SAFFLO  E RA   KE P  S  E  K LGE WK LS KEK KY   A   KA YE     Q  K G P   KK
                R

*: NFS
```

Fig. 6. Alignment of the HMG-like regions of similarity in human TCF-1; *SRY* (Sinclair *et al.*, 1990; Gubbay *et al.*, 1990); *S.pombe Mc*-mating type gene (Kelly *et al.*, 1988); *Tetrahymena* non-histone chromosomal protein LG-1 (Roth *et al.*, 1987); four domains in the human pol I transcription factor UBF (Jantzen *et al.*, 1990); and two domains in human HMG-1 (the N-terminal domain previously went unnoticed) (Wen *et al.*, 1989). The numbers in front of the amino acid sequences indicate the position of the first residue in the original protein sequences. Residues occurring at the same position in at least four of the domains are given at the bottom (consensus) and are underlined when they occur in TCF-1. O indicates aromatic residues (F, Y, W). An asterisk marks a two amino acid insertion in the N-terminal HMG domain.

and four related murine genes are described. *SRY* was identified as a gene mapping to the sex-determining region of the human and murine Y chromosome, encoding a testis-specific transcript. *SRY* was noted to contain a HMG box similar to that of *Mc* (Sinclair *et al.*, 1990; Gubbay *et al.*, 1990). The region of homology defined by our computer searches in TCF-1, Mc, LG-1 and HMG-1 aligned well with the HMG box sequences in *UBF* and in *SRY*. In fact, TCF-1 showed the highest degree of similarity to the HMG box of *SRY* (27% over 77 amino acids). In addition, we detected a second HMG box in human *HMG*-1. Figure 6 depicts the alignment of the HMG boxes of all proteins mentioned in this section.

Recombinant TCF-1 and 'T cell' TCF-1 have identical DNA-binding characteristics

To confirm the identity of the putative TCF-1 cDNA clones, the binding characteristics of recombinant TCF-1A, TCF-1B, and TCF-1C were compared with that of TCF-1 present in T cell nuclear extracts. To this end, the basic domains of the three alternative splices were subcloned into the protein A-expression vector pRIT2T (Uhlen *et al.*, 1983). After appropriate induction, the fusion proteins were purified from bacterial lysates over IgG−Sepharose and tested by gel retardation analysis. All three splice forms bound to the MW56 probe, and failed to interact with the MW56Sac probe, indicating the importance of the AACAAAG motif for binding (Figure 7A). Subsequent methylation interference footprinting (Figure 7B) proved the actual contact bases to be identical to those depicted in Figure 4 for T cell-derived TCF-1.

The TCF-1 gene is expressed in a T cell-specific fashion

As a second confirmation of identity, the cell type-specific expression of the candidate *TCF*-1 gene was analysed by Northern hybridization, using the insert of ΦTCF-1 as a probe. RNA was extracted from cell lines of the panel used in the gel retardation analyses. As shown in Figure 8, hybridizing bands of approximately 3.0 kb were observed with all T cell lines, and were absent from all other cell lines. Identical results were obtained with a full length TCF-1B probe (not shown). Thus, the presence of a signal in Northern blotting using a TCF-1 probe was completely concordant with the appearance of the TCF-1 doublet in the gel retardation assay. Taken together, the DNA-binding characteristics and mRNA expression data indicated that the cloned gene indeed encoded the *CD3-ϵ* enhancer-binding protein characterized in nuclear extracts of T lymphocytes.

Recombinant TCF-1 transactivates its cognate motif

To test the ability of TCF-1 to activate transcription, a full length version of TCF-1A was constructed in the eukaryotic expression vector pCDM7 (Aruffo and Seed, 1987) and cotransfected with the $pMW56_7$ CAT vector into COS cells. In five independent experiments, expression of TCF-1A resulted consistently in a 3- to 4-fold increase in measured CAT activity, as compared with cotransfection of $pMW56_7$ with an irrelevant vector. The outcome of a typical experiment is depicted in Table II. The observed enhancement was in the same range as that obtained with the reporter plasmids in T cells (see Table I). No effect was seen upon co-transfection of $pMW56Sac_7$ with the TCF-1A expression vector, indicating that the TCF-1A-mediated transactivation occurred through its cognate sequence. In two experiments, the dependence on cognate DNA binding was confirmed by cotransfection of the CAT reporter plasmids with an internal in-frame deletion clone of TCF-1A (TCF-1Aδ), which lacks part of the HMG box. This deletion completely abrogates DNA binding in the protein A system (not shown). Consequently, no transactivation was observed with pTCF-1Aδ (Table II). To control for transfection efficiency and survival of the reporter plasmid, quantitative Hirt extraction was performed at the time of assay on one-tenth of each cell sample. Numbers of *E.coli* colonies were counted and are indicated in brackets behind the obtained CAT values in Table II.

Discussion

The TCF-1 gene product described in this study encodes a novel type of polymerase II transcription factor. No homologies were found with such factors of other higher eukaryotes. However, computer-assisted data bank searches revealed a significant similarity of the DNA-binding domain of TCF-1 to several genes with suspected or proven DNA-binding properties. The gene most relevant to the present study is the very recently cloned polymerase I transcription factor UBF (Jantzen *et al.*, 1990). UBF controls transcription of ribosomal RNA genes by binding in a sequence-specific fashion to the Upstream Control Element of the rRNA promoter. UBF contains four domains of moderate similarity to HMG proteins. (Our sequence comparison reveals that human HMG-1 actually consists of two such

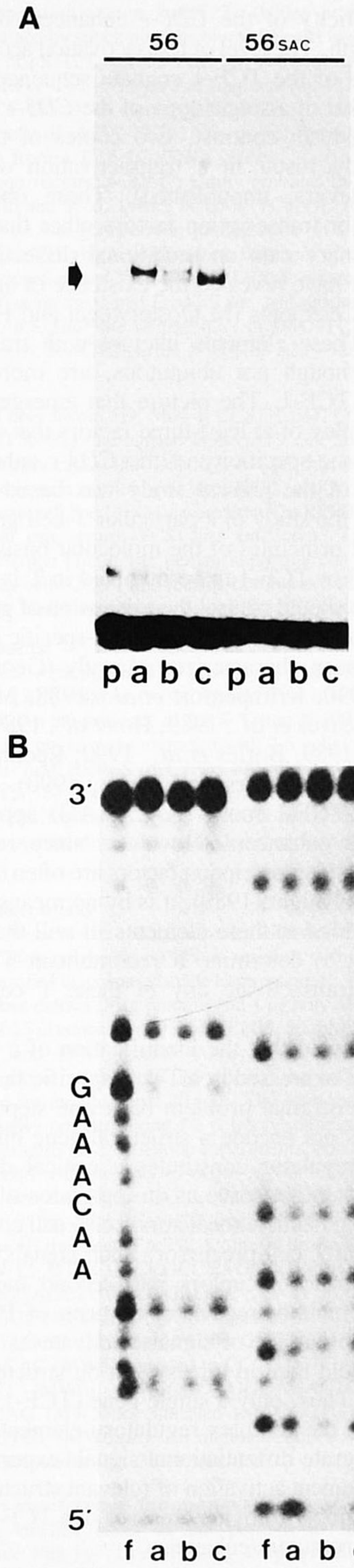


Fig. 7. All three alternative splice forms of TCF-1 recognize the same motif. The basic and alternative domains of TCF-1A/B/C were fused to the protein A-reading frame of pRIT2T. Purified fusion proteins were analysed in the gel retardation assay (**A**) and by methylation interference footprinting (**B**). p: 'non-recombinant' protein A. a: protein A/TCF-1A fusion. b: protein A/TCF-1B fusion. c: protein A/TCF-1C fusion. Gel retardation was performed with the MW56 and the MW56Sac probe. All three fusion proteins formed specific complexes with the MW56, but not with MW56Sac. The methylation interference footprints of the fusion proteins were identical to that of TCF-1 as present in T cell nuclear extracts (see Figure 4). f: free probe.

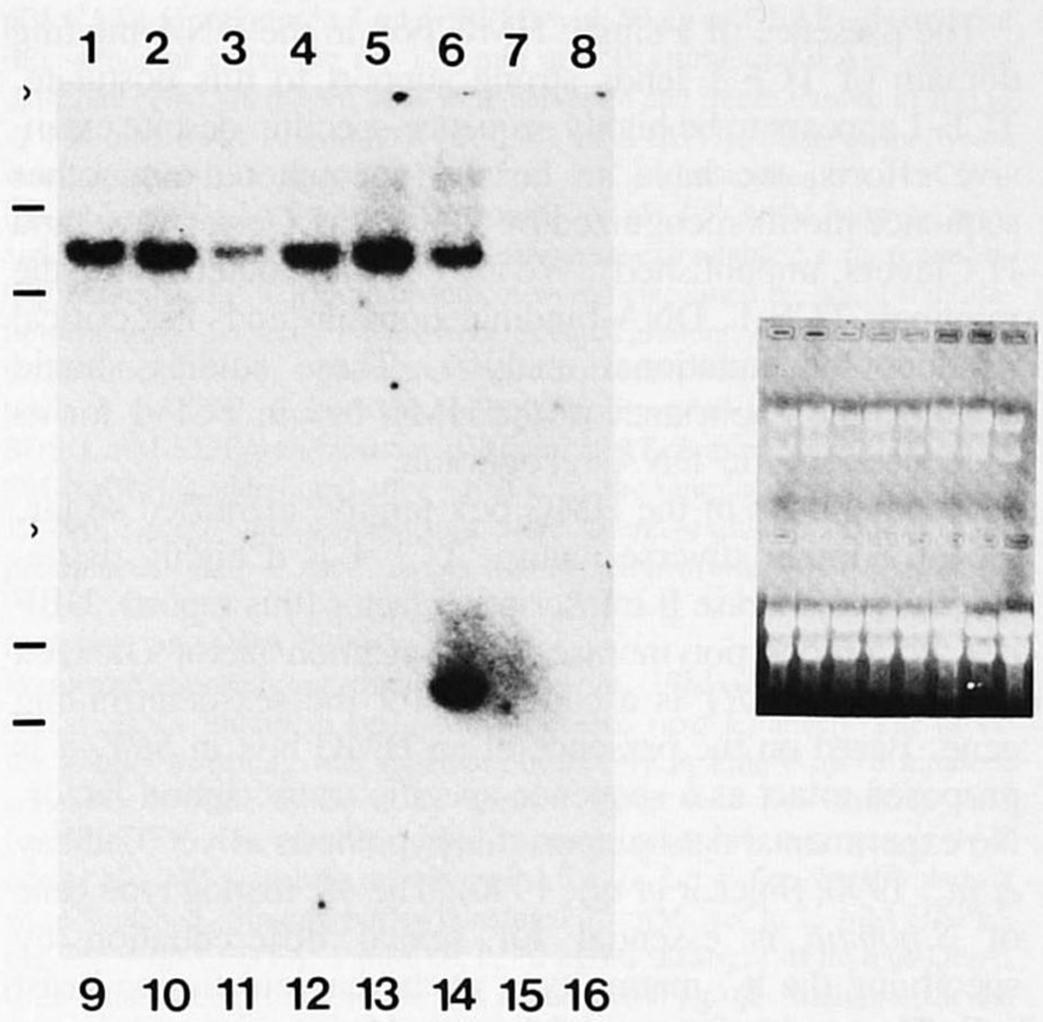


Fig. 8. Northern blot analysis of cell type-specific TCF-1 mRNA expression. Total RNA from human T cell lines (1−6 are H9, CCRF-CEM, Peer, Molt-4, CCRF-HSB, and Jurkat respectively), human non-T cell lines (7−13 are Daudi, CRL-1484, KG-1, K562, Meg-01, HeLa, and HepG2 respectively), the mouse T cell line EL-4 (14) and mouse non-T cell lines Ag8 (15) and NIH-3T3 (16) was separated on agarose, blotted and probed with the ΦTCF-1 insert. Hyphens mark the ribosomal RNA bands; arrow-heads mark the application slots. Approximately equal amounts of RNA were analysed as is evident from the ethidium bromide-stained gel (right panel).

Table II. Recombinant TCF-1 transactivates its cognate motif in non T-cells

	pCDM7-TCF-1A	pCD63	pCDM7-TCF-1Aδ
$pTK56_7$	7233 (48)	1996 (68)	2014 (49)
	6422 (38)	1705 (52)	2191 (30)
$pTK56Sac_7$	1600 (57)	1841 (29)	1820 (31)
	2122 (46)	1825 (25)	1590 (45)

COS-1 cells were transfected with CAT vectors containing seven copies of the TCF-1 motif oligonucleotide ($pTK56_7$), or containing seven copies of a mutant of the TCF-1 oligonucleotide which does not bind TCF-1 ($pTK56Sac_7$, see Table I). Cotransfected with these CAT vectors were vectors expressing TCF-1A (pCDM7-TCF-1A), the irrelevant CD63 membrane antigen (pCD63, M.Metzelaar and H.Clevers, unpublished), or a non-DNA-binding deletion clone of TCF-1A which lacks part of the HMG box (pCDM7-TCF-1Aδ). Expression of TCF-1A in COS cells consistently resulted in a 3- to 4-fold enhancement of CAT activity, provided that the CAT vector contained the TCF-1 cognate motif. The deletion clone of TCF-1A was inactive in this assay. Transfections were performed in duplicate; the results are represented as c.p.m. of extractable (butyrylated) [^{14}C]chloramphenicol. As a control for transfection efficiency, quantitative Hirt extraction was performed. The obtained *E.coli* colonies are indicated in brackets. Note that the numbers of colonies reflects the amount of (pUC18-based) CAT plasmids; the pCDM7-derived plasmids will not confer ampicillin-resistance to *E.coli* strain DH1.

boxes separated by a small linker region). The most N-terminal HMG-like domain in UBF was shown to be both necessary and sufficient for sequence-specific DNA-binding. The HMG-like protein sequence was named the HMG box, and postulated to represent a novel class of DNA-binding domain (Jantzen *et al.*, 1990).

The presence of a single HMG box in the DNA-binding domain of TCF-1 lends strong support to this postulate. TCF-1 appears to be highly sequence-specific: despite extensive efforts, we have so far not encountered any other sequence motifs recognized by TCF-1 (M.Oosterwegel and H.Clevers, unpublished). We are currently determining the minimal TCF-1 DNA-binding domain and its critical residues by mutational analysis. These studies should establish the importance of the HMG box in TCF-1 for its sequence-specific DNA recognition.

The members of the HMG box family, identified so far, are of a rather diverse nature. TCF-1 is a highly tissue-specific polymerase II transcription factor (this report). UBF is a ubiquitous polymerase I transcription factor (Jantzen *et al.*, 1990). *SRY* is a candidate for the sex-determining gene. Based on the presence of an HMG box in *SRY*, it is proposed to act as a sequence-specific transcription factor. No experimental data support this hypothesis as yet (Gubbay *et al.*, 1990; Sinclair *et al.*, 1990). The *Mc* mating type gene of *S.pombe* is essential for sexual differentiation by specifying the h^- mating type of the haploid fission yeast cell. The mode of action of the *mat-Mc* gene is not known, although it has been suggested to encode a DNA-binding protein based on the abundance of basic residues (Kelly *et al.*, 1988). *Tetrahymena* LG-1 has been studied as an HMG-1-like protein of unknown function, possibly involved in determining the macronuclear phenotype. No sequence-specific DNA-binding has been reported for LG-1 (Schulman *et al.*, 1987). Finally, the HMG-1 proteins are ubiquitous, abundant nuclear proteins capable of binding single- and double-stranded DNA in a relatively sequence-independent manner. Again, the function of these proteins is unclear, although they have been implicated in chromatin assembly (Bernues *et al.*, 1986) and transcription control (Tremethick and Molloy, 1986; Watt and Molloy, 1988).

At least three alternative splice forms are generated from the TCF-1 gene. The proteins encoded by these splice alternatives differ in their extreme C-terminus. These differences are located near, but outside, the DNA-binding domain of TCF-1, since they do not appear to affect DNA recognition. However, it cannot be excluded that the individual C-termini exert subtle effects on the DNA-binding domain proper, resulting in differences in affinity for variants of the AACAAAG motif. Experiments to test this possibility will have to await the elucidation of the spectrum of sequence motifs that can be recognized by TCF-1.

More likely however, the C-termini of TCF-1A. -B, and -C provide surfaces for protein−protein interaction with other transcription factors bound to the same DNA template, or with non-DNA binding accessory proteins (Johnson and McKnight, 1989). Due to differences in the ability to interact with other proteins, the splice forms of TCF-1 might differ in their effects on the functional status of a given regulatory DNA element. Conversely, it is conceivable that the complement of transcription factors bound to a particular regulatory element selectively allows binding of only one splice form of TCF-1. In this scheme, each particular splice form of TCF-1 would be involved in the control of an individual (set of) gene(s). Whatever the reason for the existence of the TCF-1 splice forms will turn out to be, all are expressed uniquely in T lymphocytes and are therefore probably involved in T cell-specific gene expression.

The transactivational properties of TCF-1 may account for the tissue-specificity of the *CD3-ϵ* enhancer, but do not explain its strength. The level of transactivation accomplished by seven copies of the TCF-1 cognate sequence is of the same order of that of a single copy of the *CD3-ϵ* enhancer: 5-fold to 10-fold. In contrast, two copies of the *CD3-ϵ* enhancer already result in a transactivation of at least 100-fold (H.Clevers, unpublished). These observations suggest a role for transcription factors other than TCF-1. Indeed, preliminary data on mutational dissection of the *CD3-ϵ* enhancer have revealed the existence of at least two additional DNA elements (M.Oosterwegel and H.Clevers, unpublished). These elements interact with transcription factors that, although not ubiquitous, are more broadly expressed than TCF-1. The picture that emerges predicts a complex interplay of at least three factors that dictate the strength and tissue-specificity of the *CD3-ϵ* enhancer.

The strategy of the present study was based upon the assumption that the study of a particular T cell gene would uncover general principles of the molecular basis of T cell differentiation. For TCF-1 to be involved in T lymphocyte differentiation, it should control the expression of genes other than *CD3-ϵ*. Several T cell- and lymphoid-specific regulatory elements have been characterized recently (Georgopoulos *et al.*, 1988, 1990; Krimpenfort *et al.*, 1988; McDougall *et al.*, 1988; Greaves *et al.*, 1989; Ho *et al.*, 1989; Winoto and Baltimore, 1989; Bories *et al.*, 1990; Redondo *et al.*, 1990; reviewed in Clevers and Owen, 1990). No exact matches to the TCF-1 motif (AAcAAAG) appear to be present in these enhancers. However, since recognition motifs of eukaryotic transcription factors are often degenerate (Johnson and McKnight, 1989), it is by no means excluded that TCF-1 can bind to these elements. It will therefore be of great interest to determine if recombinant TCF-1 can recognize and transactivate any of these T cell-specific enhancers.

It might be argued that the identification of a transcription factor, itself expressed in a T cell-specific fashion, has only moved the original problem back one step. The fact that TCF-1 does not encode a structural gene but rather a transcriptional regulator constitutes a critical difference, however. TCF-1 might serve as an integrator of differentiation signals (e.g. soluble mediators, cell−cell contact) that are received by a T cell precursor. Such signals would be translated by surface receptors into second messengers, finally converge onto the regulatory elements of TCF-1, and (given an appropriate set of signals) activate its transcription. TCF-1 would then in turn switch on structural genes such as *CD3-ϵ*. Thus, only a single gene (TCF-1) needs to be outfitted with the complex regulatory elements that can 'sense' and integrate diffentiational signals experienced by a cell. The subsequent activation of relevant structural genes will then result simply from the presence of a TCF-1 cognate motif in their promoters/enhancers.

Materials and methods

Plasmid constructions

CAT vectors: pBLCAT2: (Luckow and Schutz, 1987); pcTK: derived by subcloning the blunted *Cla*I−*Kpn*I fragment from the *CD3-ϵ* enhancer deletion clone pI9-c (Clevers *et al.*, 1989) into the blunted *Bam*HI site of pBLCAT2 (Figure 1); $pMW56_7$: derived by subcloning a blunted 7-mer of MW56 (see below) into blunted *Xba*I−*Bam*HI digested pBLCAT2; $pMW56Sac_7$: derived by subcloning a blunted 7-mer of MW56Sac (see below) into blunted *Xba*I−*Bam*HI digested pBLCAT2.

Protein A expression vectors: pRIT2T: (Uhlen *et al.*, 1983); pRIT2-TCF-1A: obtained by subcloning the blunted *Eco*RI insert of ΦTCF-1 into the *Sma*I site of pRIT2T; pRIT2-TCF-1B: obtained by replacing the *Pst*I fragment of pRIT2-TCF-1A by a *Pst*I fragment (bp 477−1085, see Figure 5) from a TCF-1B cDNA clone; pRIT2-TCF-1C: obtained by replacing the *Pst*I fragment of pRIT2-TCF-1A by a *Pst*I fragment (bp 477−1182, see Figure 5) from a TCF-1C cDNA clone.
COS expression vectors: pCDM7: kindly provided by Drs A.Aruffo and B.Seed (Aruffo and Seed, 1987); pCDM7-TCF-1A: A full length cDNA clone of TCF-1A was generated by joining the 5′ of a full length TCF-1B clone (*Xho*I/*Eco*47III (bp 1−620) with a fragment of the ΦTCF-1 TCF-1 *Eco*RI insert cloned into pUC19 (*Eco*47III−*Sal*I, where *Sal*I is a polylinker site of pUC19) using the shared *Eco*47III site, and subcloning this *Xho*I−*Sal*I fragment into the *Xho*I site of pMNC7; pCDM7-TCF-1Aδ: deletion of pCDM7-TCF-1A *Eco*47III-partial *Pvu*II (bp 623−716, see Figure 5b).

Cells

Human T cell lines Jurkat, H9, Molt-4, CCRF-CEM, CCRF-HSB (all TCR-α/β), Peer (TCR-γ/δ); the murine TCR-α/β T cell line EL-4; the human B lineage cell lines CRL 1484, Daudi, N-67, BS-5.2; the murine B lineage line Ag-8; the human myeloid cell lines K562, and KG-1; and the human megakaryoblastic cell line MEG-01 were grown in RPMI-1640 supplemented with 5% fetal calf serum (FCS) and antibiotics. The human cervical carcinoma cell line HeLa, the human hepatoma cell HepG2, the green monkey kidney cells COS-1 and murine fibroblasts Swiss-3T3 were grown in Dulbecco's MEM supplemented with 8% FCS and antibiotics.

Nuclear extracts

Nuclear extracts were prepared by gentle lysis of $1\times10^7-1\times10^8$ cells in STKM buffer [30% sucrose (w/v); 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 STKM, 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 r.p.m; 5 min). Protein concentration of the clear supernatant was determined and nuclear extracts were stored at −70°C.

Gel retardation assay

Fragments of the enhancer and the annealed oligonucleotides were labelled by Klenow DNA polymerase I fill in-reaction with [α-^{32}P]dCTP. 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 preincubation at room temperature, probe (10 000−20 000 c.p.m.) was added and the mixture was incubated for an additional 20 min. The samples were then electrophoresed through a non-denaturing 4% polyacrylamide gel run in 0.25×TBE at room temperature. In competition experiments, non-labelled competitor DNA was added with the poly dI-dC.
Probes used: (see also Figure 1A). L: the 83 bp *Xba*I−*Pvu*II fragment of pcTK; S: the 61 bp *Xba*I−*Eco*47III fragment of pcTK; L′: the 100 bp *Eco*47III−*Bam*HI fragment of pcTK; S′: the 74 bp *Pvu*II−*Bam*HI fragment of pcTK. MW12: GTGCCTCCGCCCAGCTGCCGCT (MW1) annealed to ACAGCGGCAGCTGGGCGGAGGC (MW2). MW34: AGCGCTCTCACACGGGCCTCCGCCC (MW3) annealed to CTGGGCGGAGGCCCGTGTGAGAGCG (MW4). MW56: GGGAGACTGAGAACAAAGCGCTCTCACAC (MW5) annealed to CCCGTGTGAGAGCGCTTTGTTCTCAGTCT (MW6). MW56Sac: GGGAGACTGAGCCGCGGTCGCTCTCACAC (MW5Sac) annealed to CCCGTGTGAGCGACCGCGGCTACAGTCT (MW6Sac). All oligonucleotides were synthesized on an Applied Biosystems Inc. 381A machine.

Methylation interference footprinting

MW56 was labelled either at the positive or the negative strand oligonucleotide with [γ-^{32}P]-ATP using T4 polynucleotide kinase. After annealing, the probes were purified over polyacrylamide. The labelled probes were partially methylated at purine residues using dimethylsulphate (Sambrook *et al.*, 1989). 100 000 c.p.m. of methylated probe was used in a five-fold 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 the G and A residues, the sequence was analysed on a 10% polyacrylamide−8 M urea sequencing gel.

CAT assays

Described in detail elsewhere (Clevers *et al.*, 1989). In short, 5×10^6 cells were transiently transfected with plasmid DNA equimolar to 10 μg of pBLCAT2 for 60 min in 3 ml of RPMI with 50 μg of DEAE−dextran/ml (non-adherent cells); or for 120 min at 250 μg/ml of DEAE−dextran (adherent cells). 48 h later, cells were harvested and freeze-thawed in 100 μl of 100 mM NaCl/10 mM Tris pH 7.4/1 mM EDTA. 50 μl of the lysate was added to 125 μl of CAT cocktail [^{14}C]chloramphenicol 1 μCi/ml (60 mCi/mmol); 2.5% glycerol; 250 mM Tris pH 7.5; 3 mM butyryl-CoA), and incubated for 2 h at 37°C. Pristane/xylene-extractable c.p.m. representing butyrylated [^{14}C]chloramphenicol were determined by liquid scintillation counting. To verify the presence of equal amounts of reporter plasmid at the time of assay, a quantitative Hirt extraction was performed on one-tenth of the cells. Cells were pelleted, resuspended in 0.2 ml of 0.6% SDS/1 mM EDTA and incubated 20 min at RT. Subsequently 0.05 ml of 5M NaCl was added, and after 6 h at 4°C, the samples were spun (5 min 15 000 r.p.m), the supernatants were phenol-extracted and ethanol-precipitated. Half of each sample was then transformed into *E.coli*.

Isolation of cDNA clones

Performed essentially according to Vinson *et al.* (1988). An oligo(dT) primed Jurkat cDNA library in λgt11 was purchased from Clontech. The probe for affinity screening was generated by briefly ligating 6 μg of kinased, annealed MW56, followed by Klenow fill-in with [α-^{32}P]dCTP and [α-^{32}P]dGTP. The resulting probe mixture had an average size of 4−8 copies of MW56; specific activity was 2.5×10^7 c.p.m./μg. 2×10^6 plaques were screened. Binding buffer contained 75 mM NaCl; 25 mM HEPES pH 7.9; 1 mM EDTA; 1 mM DTT; 10 μg/ml sheared calf thymus DNA; 0.25% non-fat dry milk. Probe was added to 100 ng/ml. Additional clones were isolated from the Jurkat library and from an HPB-ALL library in pCDM7 by standard hybridization screening with the insert of ΦTCF (Sambrook *et al.*, 1989).

DNA sequencing

Sequencing was performed on double stranded DNA templates according to the method of Sanger *et al.* (1977) using T7 DNA polymerase (Pharmacia). 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 in RNAzol according to the manufacturer's procedures (Cinna/Biotecx), followed by phenol−chlorofrom extraction and 2-propanol precipitation. 10 μg of total RNA was run for Northern analysis on 1% agarose containing 6% formaldehyde. RNA transferred to nitrocellulose was hydridized with the insert of ΦTCF labelled by random oligonucleotide priming according to standard procedures (Sambrook *et al.*, 1989).

Expression of TCF-1A, -B and -1C in E.coli

E.coli strain N4830-1 was used as the host for expression of TCF-1 using the heat-inducible protein A expression vector pRIT2T (Uhlen *et al.*, 1983). 400 ml of LB containing 100 μg/ml ampicillin was inoculated with 25 ml of an overnight culture, and grown at 30°C to $A_{600}\sim0.9$. 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 5000 r.p.m. for 10 min, and the pellet was resuspended in 80 ml of 50 mM Tris pH 7.6, 150 mM NaCl, 0.05% Tween, 1 mM PMSF. The bacteria were lysed by sonification (4×4 min) on ice. Cell fragments were removed by centrifugation (30 min, 15 000*g*). The fusion protein were purified over IgG-Sepharose according to the manufacturer's instructions (Pharmacia) and stored at −70°C.

Acknowledgements

The authors wish to thank Drs A.Aruffo and B.Seed for providing pMNC7 and the HPB-ALL cDNA library; Dr E.Rozemuller for assistance with computer searches; Dr M.Tilanus for help with the pRIT2T system; Drs S.Verbeek, J.Deschamps, F.Meijlink, and T.Logtenberg for critically reading the manuscript; A.Verdonk for secretarial help, and Professor P.v.d.Vliet and the members of the laboratory for helpful discussions and support.

References

Aruffo,A. and Seed,B. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 3365−3369.
Atchison,M.L. (1988) *Annu. Rev. Cell. Biol*, **4**, 127−153.
Bernues,J., Espel,E. and Querol,E. (1986) *Biochim. Biophys. Acta*, **866**, 242−251.

Bodner,M., Castrillo,J.-L., Theill,L.E., Deerinck,T., Ellisman,M. and Karin,M. (1988) *Cell*, **55**, 505–518.
Bories,J.C., Loiseau,P., d'Auriol,L., Gontier,C., Bensussan,A., Degos,L. and Sigaux,F. (1990) *J. Exp. Med.*, **171**, 75–83.
Clevers,H.C., Dunlap,S., Wileman,T.E. and Terhorst,C. (1988a) *Proc. Natl. Acad. Sci. USA.*, **85**, 8156–8160.
Clevers,H., Alarcon,B., Wileman,T. and Terhorst,C. (1988b) *Annu. Rev. Immunol.*, **6**, 629–662.
Clevers,H., Lonberg,N., Dunlap,S., Lacy,E. and Terhorst,C. (1989) *EMBO J.*, **8**, 2527–2535.
Clevers,H. and Owen,M. (1990) *Immunology Today*, in press.
Davis,R.L., Weintraub,H. and Lassar,A.B. (1987) *Cell*, **51**, 987–1000.
Evans,T. and Felsenfeld,G. (1989) *Cell*, **58**, 877–885.
Frain,M., Swart,G., Monaci,P., Nicosia,A., Stampfli,S., Frank,R. and Cortese,R. (1989) *Cell*, **59**, 145–157.
Georgopoulos,K., Van den Elsen,P., Bier,E., Maxam,A. and Terhorst,C. (1988) *EMBO J.*, **7**, 2401–2408.
Georgopoulos,K., Galson,D. and Terhorst,C. (1990) *EMBO J.*, **9**, 109–115.
Greaves,D.R., Wilson,F.D., Lang,G. and Kioussis,D. (1989) *Cell*, **56**, 979–986.
Gubbay,J., Collignon,J., Koopman,P., Capel,B., Economou,A., Munsterberg,A., Vivian,N., Goodfellow,P. and Lovell-Badge,R. (1990) *Nature*, **346**, 245–250.
Ho,I.-C, Yang,L.-H., Morle,G. and Leiden,J.M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 6714–6718.
Ingraham,H.A., Chen,R., Mangalam,H.J., Elsholtz,H.P., Flynn,S.E., Lin,C.R., Simmons,D.M., Swanson,L. and Rosenfeld,M.G. (1988) *Cell*, **55**, 519–529.
Jantzen,H.-M., Admon,A., Bell,S.P. and Tjian,R. (1990) *Nature*, **344**, 830–836.
Johnson,P.F. and McKnight,S.L. (1989) *Annu. Rev. Biochem.*, **58**, 799–839.
Kelly,M., Burke,J., Smith,M., Klar,A. and Beach,D. (1988) *EMBO J.*, **7**, 1537–1547.
Kozak,M. (1984) *Nucleic Acids Res.*, **12**, 857–872.
Krimpenfort,P., de Jong,R., Uematsu,Y., Dembic,Z., Ryser,S., von Boehmer,H., Steinmetz,M. and Berns,A. (1988) *EMBO J.*, **7**, 745–750.
Lassar,A.B., Buskin,J.N., Lockshon,D., Davis,R.L., Apone,S., Hauschka,S.D. and Weintraub,H. (1989) *Cell*, **58**, 823–831.
Lipman,D.J. and Pearson,W.R. (1985) *Science*, **227**, 1435–1441.
Luckow,B. and Schutz,G. (1987) Nucleic Acids Res., **15**, 5490.
Maniatis,T., Goodbourne,S. and Fischer,J.A. (1987) *Science*, **236**, 1237–1245.
McDougall,S., Peterson,C.L. and Calame,K. (1988) *Science*, **241**, 205–208.
Mermod,N., O'Neill,E.A., Kelly,T.J. and Tjian,R. (1989) *Cell*, **58**, 741–753.
Mueller,C.R., Maire,P. and Schibler,U. (1990) *Cell*, **61**, 279–291.
Mueller,M.M., Ruppert,S., Schaffner,W. and Matthias,P. (1988) *Nature*, **336**, 544–551.
Redondo,J.M., Hata,S., Brocklehurst,C. and Krangle,M.S. (1990) *Science*, in press.
Roth,S.Y., Schulman,I.G. Cook,R.G. and Allis,C.D. (1987) *Nucleic Acids Res.*, **15**, 8112.
Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual. 2nd edition*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Scheidereit,C., Cromlish,J.A., Gerster,T., Kawakami,K., Balmaceda,C-G., Currie,A. and Roeder,R.G. (1988) *Nature*, **336**, 551–557.
Schulman,I.G., Cook,R.G., Richman,R. and Allis,C.D. (1987) *J. Cell Biol.*, **104**, 1485–1494.
Scott,M.P. and Carroll,S.B. (1987) *Cell*, **51**, 689–698.
Sinclair,A.H., Berta,P., Palmer,M.S., Hawkins,J.R., Griffiths,B.L., Smith,M.J., Foster,J.W., Frischau,A.-M., Lovell-Badge,R. and Goodfellow,P.N. (1990) *Nature*, **346**, 240–244.
Singh,H., LeBowitz,J.H., Sharp,P.A. and Baltimore,D. (1988) *Cell*, **52**, 415–423.
Tremethick,D.J. and Molloy,P.L. (1986) *J. Biol. Chem.*, **261**, 6986–6992.
Tsai,S-F., Martin,D.I., Zon,L.I., D'Andrea,A.D., Wong,G.G. and Orkin,S.H. (1989) *Nature*, **339**, 446–451.
Tunnacliffe,A., Olsson,C., Buluwela,C. and Rabbits,T.H. (1988) *Eur. J. Immunol.*, **18**, 1639–1642.
Uhlen,M., Nilsson,B., Guss,B., Lindberg,M., Gatenbeck,S. and Philipson,L. (1983) *Gene*, **23**, 369–378.
Watt,F. and Molloy,P.L. (1988) *Nucleic Acids Res.*, **16**, 1471–1486.
Wen,L., Huang,J.-K, Johnson,B.H. and Reeck,G.R. (1989) *Nucleic Acids Res.*, **17**, 1197–1214.
Weintraub,H., Tapscott,S.J., Davis,R.L., Thayer,M.J., Adam,M.K., Lassar,A.B. and Miller,A.D. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 5434–5438.
Winoto,A. and Baltimore,D. (1989) *EMBO J.*, **8**, 729–733.
Vinson,C.R., LaMarco,K.L., Johnson,P.F., Landschultz,W.H. and McKnight,S.L. (1988) *Genes Dev.*, **2**, 801–806.

Received on October 1, 1990; revised on November 7, 1990