MOLECULAR AND CELLULAR BIOLOGY, Mar. 1996, p. 745–752 Vol. 16, No. 3
0270-7306/96/$04.00+0

Extensive Alternative Splicing and Dual Promoter Usage Generate Tcf-1 Protein Isoforms with Differential Transcription Control Properties

MARC VAN DE WETERING, JAN CASTROP, VLADIMIR KORINEK, AND HANS CLEVERS*

Department of Immunology, University Hospital Utrecht, 3508 GA Utrecht, The Netherlands

Received 11 July 1995/Returned for modification 11 September 1995/Accepted 28 November 1995

Previously, we reported the isolation of cDNA clones representing four alternative splice forms of TCF-1, a T-cell-specific transcription factor. In the present study, Western blotting (immunoblotting) yielded a multitude of TCF-1 proteins ranging from 25 to 55 kDa, a pattern not simply explained from the known splice alternatives. Subsequent cDNA cloning, PCR amplification, and analysis by rapid amplification of 5′ cDNA ends revealed (i) the presence of an alternative upstream promoter, which extended the known N terminus by 116 amino acids, (ii) the presence of four alternative exons, and (iii) the existence of a second reading frame in the last exon encoding an extended C terminus. Inclusion of the extended N terminus into the originally reported protein resulted in a striking similarity to the lymphoid factor Lef-1. Several of the TCF-1 isoforms, although less potent, mimicked Lef-1 in transactivating transcription through the T-cell receptor α-chain (TCR-α) enhancer. These data provide a molecular basis for the complexity of the expressed TCF-1 proteins and establish the existence of functional differences between these isoforms. Furthermore, the functional redundancy between Tcf-1 and Lef-1 explains the apparently normal TCR-α expression in single *Tcf-1* or *Lef-1* knockout mice despite the firm in vitro evidence for the importance of the Tcf/Lef site in the TCR-α enhancer.

We previously identified and cloned the T-lymphocyte-specific transcription factor TCF-1 (15, 22) by virtue of its affinity for the sequence motif AACAAAG in the enhancer of a structural T-cell gene, CD3-ε. We recently reported on a gene disruption experiment which revealed that the *TCF-1* gene is essential for early T-cell differentiation (28). TCF-1 belongs to a family of proteins that utilize a region with homology to high-mobility-group (HMG) I proteins for sequence-specific DNA binding. This region has been termed the HMG box (10, 11). Solution structures have recently been proposed for Lef-1 and Sry HMG boxes complexed to DNA (13, 31). In addition, we have solved the three-dimensional solution structure for the related Sox-4 HMG box (26).

Human LEF-1, which is closely related to TCF-1, was originally identified as a T-cell-specific protein binding to the TTCAAAG motif in the T-cell receptor α-chain (TCR-α) enhancer (29). Murine Lef-1 was simultaneously cloned from pre-B lymphocytes by a subtraction strategy and was shown to bind to the same TCR-α motif (19). In postnatal mice, Tcf-1 is uniquely expressed by T-lineage lymphocytes, whereas Lef-1 is expressed by T cells and by pre-B cells. By contrast, Tcf-1 and Lef-1 exhibit a complex and largely overlapping expression pattern in several nonhemopoietic tissues during murine embryogenesis (16). Tcf-1 and Lef-1 derive from a recent gene duplication event, as evidenced by the cloning of a single chicken homolog (3). The HMG boxes of Tcf-1 and Lef-1 are virtually identical and likely display indistinguishable DNA-binding specificities. Tcf-1 and Lef-1 binding motifs have been found in several other lymphoid cell-specific enhancers, including those of the TCR-β, TCR-δ, and CD4 genes (5, 12). Surprisingly, Tcf-1 and Lef-1 do not transactivate transcription from a reporter chloramphenicol acetyltransferase (CAT) gene that carries multiple copies of their minimal binding motifs (19, 22, 29). Nevertheless, several studies have indicated that the Lef-1 site in the TCR-α enhancer is functionally important and that Lef-1 can transactivate transcription of a reporter gene through this site (1, 8, 9). This transactivation is context specific, depending for its full activity on the simultaneous presence of three other factors (ATF/CREB, core-binding factor/PEBP2α, and Ets-1) at defined flanking positions (9). The HMG box of Lef-1 in isolation retains some context-specific transactivation properties in the TCR-α model, as does the HMG box of the distantly related Sry protein (9). Thus, part of the Lef-1 effect apparently relates to its ability to strongly bend DNA (7), a shared characteristic of sequence-specific HMG box proteins (6). In an unrelated transcription model, it was shown by Sheridan et al. that Lef-1 controls activity of the human immunodeficiency virus enhancer in a chromatin-dependent fashion (17). Specifically, Lef-1 was shown to counteract the repressive effect of chromatin assembly on the human immunodeficiency virus enhancer. No candidate target genes for Tcf-1 have been identified to date.

We originally reported that mRNA transcribed from the human *TCF-1* gene was subject to alternative splicing in the region encoding the extreme C terminus of the protein, giving rise to four mRNA molecules: TCF-1A to TCF-1D (22, 23). The predicted molecular masses of these protein isoforms are about 30 kDa. Two observations prompted us to reevaluate the complexity of transcripts originating from the *TCF-1* gene. (i) A minigene encoding these four isoforms was correctly expressed but did not complement the *Tcf-1* knockout phenotype (27, 28). (ii) Western blotting (immunoblotting) performed with an anti-TCF-1 monoclonal antibody revealed a multitude of bands between 25 and 55 kDa (this report). Here we report on the identification and functional characteristics of novel alternative transcripts emanating from the *TCF-1* locus,

* Corresponding author. Mailing address: Department of Immunology, University Hospital Utrecht, P.O. Box 85500, 3508 GA Utrecht, The Netherlands.

some of which are remarkably similar in overall structure to Lef-1.

MATERIALS AND METHODS

Isolation of *TCF-1* cDNA clones. Standard techniques were used for screening a human HPB-ALL library in pCDNA (10a) and a mouse thymus cDNA library in Lambda-zap (Stratagene). Probes encoding human exon IX and mouse 5′ sequences were generated by PCR from the relevant cDNA clones (15, 22). Oligonucleotides were obtained from Isogen (Amsterdam, The Netherlands).

Constructs. Hybrid TCF-1 constructs were generated by cloning a mouse PCR fragment, encompassing amino acids 1 to 130, into a *Bam*HI-*Apa*LI-digested human TCF-1B clone. Oligonucleotides used were AGTGGATCCAGCGCAC CATGCCGCAGCTG and AGTGGTGCACTCCGGGGCCTT. TCF-LEF was constructed by PCR of the region encoding amino acids 1 to 118 of pCDNA-p^{45}TCF-1B, using vector T7 primer and oligonucleotide GTAGAGGCCTCCG GTGCACTCCGGGGC. This fragment was digested with *Hin*dIII and *Stu*I and cloned into *Hin*dIII-*Stu*I-digested pCDNA-hLEF-1.

δLEF-1 was constructed by PCR of the region encoding amino acids 100 to 214 and addition of an optimal Kozak sequence, using the primers GGGAAGCT TGCCACCATGTACAACAAG and CTTGCCAGCCAAGAGGTGGGGTG. This fragment was digested with *Hin*dIII and *Eco*RV and cloned into *Hin*dIII-*Eco*RV-digested pCDNA-hLEF-1.

RACE. Total RNA from human thymus was prepared in RNAzol as instructed by the manufacturer (Cinna/Biotecx); the preparation was then subjected to phenol-chloroform extraction and 2-propanol precipitation. Rapid identification of 5′ cDNA ends (RACE) analysis was performed as follows. Five micrograms of RNA was used to generate cDNA, using avian myeloblastosis virus reverse transcriptase and TCF-1-specific primers (Isogen). After separation of the reaction product from the excess of primers, cDNAs were tailed with dATP and terminal deoxynucleotidyltransferase; 1/100 of the tailed products was used in a standard PCR using tail- and TCF-1-specific primers.

Cell culture and transfection. The murine surface immunoglobulin G2A-positive B-cell line AZU2 was grown in RPMI 1640 supplemented with 10% fetal calf serum and antibiotics. AZU2 cells were transfected by electroporation as described previously (24).

COS cells were grown in Dulbecco modified Eagle medium supplemented with 5% fetal calf serum and antibiotics. Transient transfections of COS cells with expression plasmids derived from pCDNA (Invitrogen) were performed by incubating the cells in suspension for 120 min with 1 μg of plasmid in 3 ml of Dulbecco modified Eagle medium with 250 μg of DEAE-dextran ml^{-1}.

Western blotting (immunoblotting). About 3×10^5 cells were used as a protein sample. The cells were prepared as protein samples by washing them with phosphate-buffered saline (PBS) and resuspending them in 30 μl of sample buffer. After sonification for 20 s and boiling for 5 min, the protein samples were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Gels were blotted to nitrocellulose membranes (Schleicher & Schuell, Keene, N.H.). The membranes were then blocked with 5% milk in washing buffer (PBS, 0.1% Tween 20) for 2 h, incubated with the anti-TCF-1 monoclonal antibody (culture supernatant) for 2 h, washed for 15 min in washing buffer, and incubated with peroxidase-conjugated rabbit anti-mouse immunoglobulin (diluted 1:2,000) for 2 h. The specific antibody signals were detected by using enhanced chemiluminescence (ECL kit; Amersham). For the quantification of TCF-1 protein in Jurkat T cells, a sample equaling 2×10^5 cells was run and compared with standard amounts of recombinant TCF-1 protein (4). After detection by enhanced chemiluminescence (ECL kit; Amersham), the bands were scanned by densitometry and the amount of TCF-1 present in Jurkat cells was determined. From this, the number of molecules per cell could be calculated.

In vitro translation and gel retardation. The TCF-1 proteins used in the gel retardation assay were produced by using the Promega SP6 TnT rabbit reticulocyte lysate coupled transcription-translation system according to the manufacturer's protocol. To verify that proteins were produced in similar quantities, 1 μl of the reaction mixture was Western blotted and stained with the anti-TCF-1 monoclonal antibody (see above). Gel retardation experiments were performed as described earlier (22). In short, 1 μl of the transcription-translation mixture was incubated with 100 ng of poly(dI-dC) and 50,000 cpm of either oligonucleotide MW56, containing a TCF-1 binding site, or MW56-Sac, which does not bind TCF-1 (22). Bound and free probe were separated on a 5% polyacrylamide gel.

CAT assays. CAT assays were performed as described earlier (22). In short, 3×10^6 cells were transiently transfected with 2.5 μg of TCR-α enhancer CAT construct (8) and 75, 750, or 7,500 ng of expression vector by electroporation. The total amount of DNA was kept constant at 10 μg by adding a nonrelevant plasmid. Forty-eight hours 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 50 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 coenzyme A), and the mixture was incubated for 2 h at 37°C. Pristane-xylene-extractable counts per minute representing butyrylated [^{14}C]chloramphenicol was determined by liquid scintillation counting.

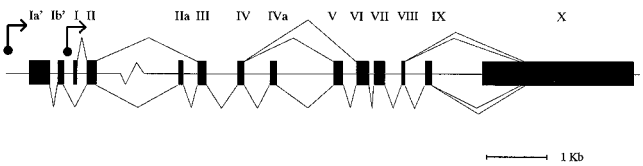

FIG. 1. Structure of the *TCF-1* gene. Exons are represented by black boxes. Promoters are depicted by arrows. The downstream *TCF-1* promoter overlaps exons Ia′ and Ib′. The predominant modes of splicing are indicated by lines.

Nucleotide sequence accession number. The sequence reported has been assigned GenBank accession number X63901.

RESULTS

To aid the clarity of this discussion, Fig. 1 gives an updated gene structure of *TCF-1*, generated by the compilation of earlier information (15, 22, 23) with data from the current study (see below).

Western blotting of human TCF-1. We have recently generated a monoclonal antibody against human TCF-1 (4). The epitope maps to a peptide sequence encoded by the constitutive exon VI. The antibody allows biochemical analyses of TCF-1 proteins of T cells. Interestingly, we have found that TCF-1 protein is absent from standard nuclear extracts, as determined by gel retardation and Western blotting. However, it is readily visualized by nuclear staining and by Western blotting of whole cells or isolated nuclei (2). We conclude that the TCF-1 proteins are intimately associated noncovalently with components of the nuclear matrix. The nature of this interaction is the subject of our ongoing research.

Western blotting performed on human thymocytes and Jurkat T cells resulted in a multitude of bands with deduced molecular masses of 25 to 55 kDa in SDS-PAGE (Fig. 2A). Identical patterns were observed when protein samples were reduced by β-mercaptoethanol, excluding the existence of disulfide-linked dimers (not shown). In addition, dephosphorylation of the samples prior to electrophoresis did not influence the banding pattern. To determine protein levels of the prominent TCF-1 isoforms in Jurkat cells, we compared immunoreactive TCF-1 in 10^6 Jurkat cells with a standard series of purified recombinant TCF-1 (Fig. 2B). Scanning of the individual bands revealed that the amounts of immunoreactive material present in the three major bands (a [P^{45}TCF-1B], b [P^{42}TCF-1B] and c [P^{33}TCF-1B]; see below) in Fig. 2B were about 60, 90, and 90 ng, respectively, equivalent to an estimated total of 3.9×10^6 molecules of TCF-1 per Jurkat T cell.

Since we could not explain the large sizes of some of the TCF-1 isoforms, we analyzed the *TCF-1* cDNA sequence for previously unidentified reading frames. We thus noted a hypothetical splice form which encoded an extended C terminus. This splice form included exon VIII, spliced to the alternative exon IX, which was spliced in turn to the 3′-most of the three acceptor sites in exon X. The resulting frameshift opened a novel reading frame in exon X. The C terminus encoded by this alternative reading frame was termed TCF-1E. (The previously identified, overlapping reading frame encoding TCF-1B results from exon VIII spliced directly to the same acceptor site in exon X.) Significantly, the reading frames of TCF-1B (15, 22) and TCF-1E were both virtually completely conserved between human and mouse sequences (Fig. 3A).

The expression of this hypothetical C-terminal reading frame was confirmed by screening a cDNA library from the human T-cell line HPB-ALL (which previously served to define TCF-1A to -D) with a probe specific for the alternative exon IX. Of

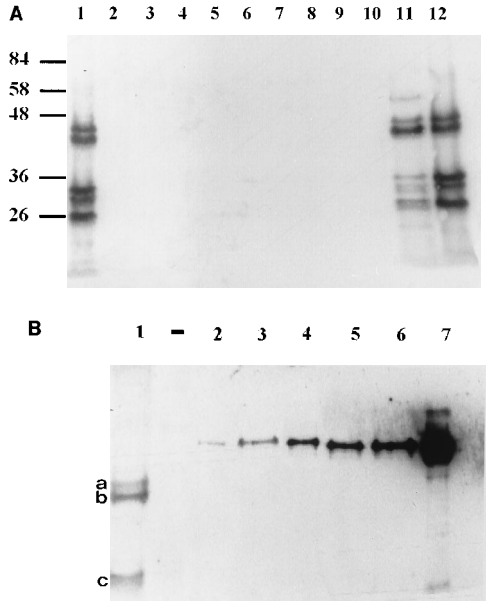

FIG. 2. Western blot analysis of TCF-1 expression in T-lineage cells. (A) The anti-TCF-1 monoclonal antibody (23) recognizes multiple bands ranging from 25 to 55 kDa in human T-cell extracts, which are absent from non-T cells. Lane 1, thymocytes. Non-T cells are in lanes 2 (Nalm-6), 3 (BJAB), 4 (Daudi), 5 (APD), 6 (U266), 7 (HL-60), 8 (K562), 9 (KG-1), and 10 (HepG2). T cells are in lanes 11 (Jurkat) and 12 (thymocytes). (B) Determination of TCF-1 protein levels in Jurkat T cells. The predominant TCF-1 isoforms are marked a, b, and c, representing P^{45}TCF-1B, P^{42}TCF-1B, and P^{33}TCF-1B, respectively. Lane 1, Jurkat T cells. Lanes 2 to 7 contain set amounts of recombinant TCF-1–maltose-binding protein fusion protein (23): lane 2, 5 ng; lane 3, 10 ng; lane 4, 20 ng; lane 5, 35 ng; lane 6, 50 ng; lane 7, 150 ng. Sizes are indicated in kilodaltons.

15 independent clones analyzed, 2 were identified as TCF-1C (exon IX spliced to the middle acceptor in exon X), whereas 13 encoded genuine TCF-1E C termini. The frequency of the TCF-1E splice form in the HPB-ALL library was estimated to be less than 5%. The vast majority of the C termini in the HPB-ALL library were of the TCF-1B type (not shown).

In the course of analyzing the putative TCF-1E clones, we noted the presence of three alternatively utilized exons. (i) An alternative exon IVA, identical at the protein level to the murine Tcf-1 IVA sequence (15), occurred in about 50% of the cDNAs. Significantly, the human *LEF-1* gene includes an alternative exon homologous to IVA at the corresponding position. Homologous sequences are also present in murine Lef-1 (19) and chicken Tcf-1 (3) (Figure 3B). (ii) Exon V was infrequently absent, resulting in an in-frame deletion of 40 amino acids. This particular splicing alternative was previously seen in the Tcf-$1^{\delta v}$ knockout mice, in which it partially rescues the knockout phenotype (28). (iii) Furthermore, we found an alternative exon IIA, which contained two in-frame stop codons (Fig. 3C). The conceptual translation product of exon IIA-containing mRNAs was a 155-amino-acid, non-DNA-binding peptide.

Expression of all these human isoforms in COS cells and subsequent Western analysis revealed that the longest of the reading frames, TCF-1E with the inclusion of the alternative exon IVA, was still significantly smaller than the predominant band observed in Jurkat T cells or in human thymocytes (Fig. 4). Unexpectedly, inspection of the published murine 5′ untranslated sequence revealed the existence of a reading frame with striking similarity to N-terminal sequences of Lef-1 and of chicken Tcf-1. The comparison indicated that the murine cDNA clone was incomplete and started at codon 55 (Fig. 3B). Importantly, the murine 5′ sequence deviated completely from exon I of human TCF-1 but was highly homologous to its human counterpart from exon II onward. To investigate whether alternative transcripts existed in human cells, we screened the HPB-ALL cDNA library with a short probe representing the unique 5′ region of the murine cDNA clone. We identified multiple clones that contained sequences homologous to the murine 5′ end, spliced to exon II. By RACE, these sequences were also found in freshly isolated human thymocytes (not shown). Unfortunately, a strong stop for reverse transcriptase was consistently encountered around amino acid position 80 (Fig. 3B). We therefore screened a murine cDNA library with the 5′ probe. Of 12 sequenced clones, one extended upstream of the expected novel translation initiation site; all others again terminated in the region of high CG content approximately 80 codons downstream of the presumed initiator methionine. Alignment of the complete N terminus of TCF-1 with Lef-1 and chicken Tcf-1 confirmed the expected sequence conservation (Fig. 3B).

All novel sequences were present on the genomic phages previously used to resolve the human *TCF-1* gene structure (23). This allowed us to draw a (presumably) complete map of the *TCF-1* gene (Fig. 1). The novel 5′ sequences could be assigned to two exons, termed exon Ia′ and exon Ib′. As these two exons were located directly upstream from the previously reported transcription start site (23), the two promoters mapped very closely together. The CpG island that we originally identified as the *TCF-1* promoter (23) started within the first exon of the upstream promoter and included the second exon (Fig. 1). Nevertheless, the previously published functional promoter data (23) reflect properties of the downstream promoter only, as the upstream promoter was not included in the tested fragment.

Expression in COS cells of several of the human splice forms with the extended N terminus and subsequent Western blotting revealed proteins of up to 60 kDa (Fig. 4). Comparison with extracts from thymocytes indicated that the predominant in vivo isoforms of TCF-1 most likely contained the extended N terminus and the TCF-1B C terminus, with or without exon IVA. In order to introduce a unequivocal nomenclature, we propose to name the TCF-1 isoforms by their molecular weights (which discriminates which N terminus is present and whether exon IVA is being used), in combination with TCF-1A to -E to distinguish the C termini. Figures 4 and 5 give eight of the predominant isoforms.

Lef-1 has been shown to transactivate transcription in a context-specific fashion through a small fragment from the T-cell-specific TCR-α enhancer (1, 8, 9). We have recently identified a murine B-cell line in which cotransfection of a Lef-1 expression construct enhances the activity of the TCR-α enhancer up to 100-fold (20). This effect is an order of magnitude higher than that reported for various other non-T-cell lines (1, 8, 12, 20). In this B-cell line, an optimal constellation of transcription factors preexists which is apparently silent by itself but strongly supports TCR-α enhancer activity upon transfection of Lef-1. The identity of these factors is the subject of ongoing research. Nevertheless, this AZU2 cell line allowed a highly sensitive assay system with which to evaluate transactivation properties of the various TCF-1 isoforms. Eight of the most prominent human TCF-1 isoforms, cloned into an appropriate expression vector, were separately cotransfected with a TCR-α enhancer/c-*fos* promoter-driven CAT reporter gene construct (8). The four isoforms with short N termini all moderately transactivated transcription through the TCR-α enhancer (Fig. 5A). The four isoforms with long N termini did

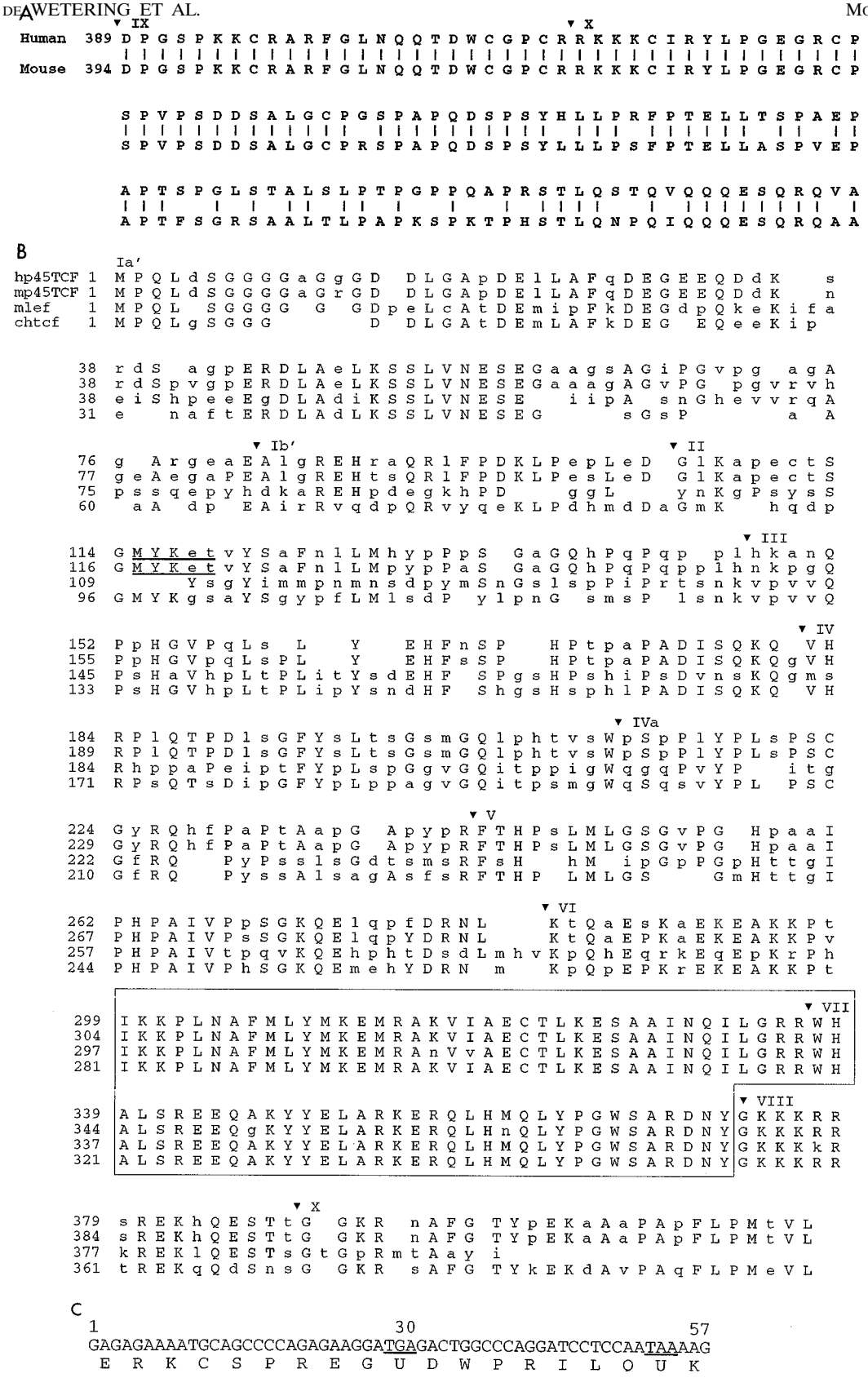

FIG. 3. Sequences of TCF-1 isoforms. (A) Alignment of the predicted amino acid sequences of the C termini of human and murine TCF-1E. A closed triangle indicates the exon boundaries. Exon numbers are indicated in roman numerals. (B) Sequence alignment of human P^{45}TCF-1B (hp45TCF), mouse P^{45}Tcf-1B (mp45TCF), mouse Lef-1 (mlef), and chicken Tcf-1 (chtcf). The DNA-binding HMG box domain is boxed. Note the high conservation between TCF-1, mouse Lef-1, and chicken Tcf-1 in the N-terminal region. The first five amino acids of the shorter TCF-1 proteins are underlined. Exon numbers and boundaries are indicated as in panel A. (C) Nucleic and amino acid sequences of the alternative exon IIa as found in human TCF-1. Insertion of this exon between exon II and III will result in the production of a 16-kDa protein lacking DNA-binding properties.

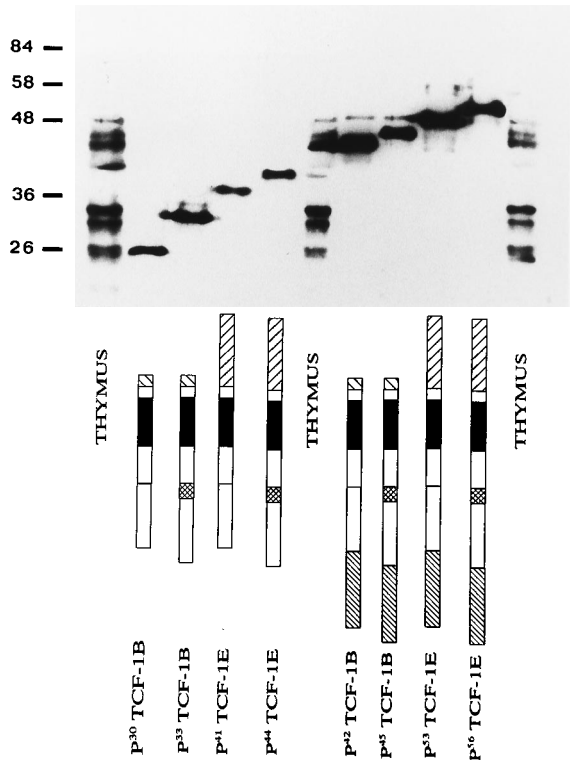

FIG. 4. Western blotting performed on COS cells transfected with cDNAs encoding different protein isoforms. The transfected TCF-1 isoforms are schematically represented under the Western blot. A cross-hatched box represents the alternative exon IVa; the densely striped boxed indicates the long N terminus; short and long C termini represent TCF-1B and TCF-1E, respectively (see also the legend to Fig. 5). "Thymus" indicates extracts from human thymocytes. Note that the apparent mobilities (indicated in kilodaltons) of some of the individual proteins appears anomalous. Also note that the strongest bands in thymocytes coincide with the various TCF-1B isoforms.

not affect activity of the reporter construct to any significant extent. This pattern was consistently seen over a range of concentrations of the expression plasmids. The inactivity of the longer forms was not due to ineffective expression: the translation initiation sequence of the longer N terminus (gcacc ATGc) adheres considerably better to the Kozak consensus than that of the shorter form (agcggcATGt). In line with this, the expression constructs with the longer forms were expressed to severalfold-higher levels than those encoding the shorter forms, as assessed by Western blotting and nuclear staining of transfected cells (not shown).

As the HMG boxes of Lef-1 and Sry in isolation transactivate transcription in the TCR-α enhancer assay (9), we tested whether this also held true for the TCF-1 HMG box. Indeed, as demonstrated in Fig. 5B, coexpression of the TCF-1 HMG box with the TCR-α–CAT vector resulted in transactivation. Figure 5B also demonstrates that chicken Tcf-1 moderately transactivated transcription through the TCR-α enhancer.

To shed more light on the basis for the functional differences between short and long TCF-1 isoforms, we performed several experiments. In vitro transcription-translation revealed that DNA-binding affinities did not significantly differ between P^{33}TCF-1 and P^{45}TCF-1, indicating that the N terminus did not directly interfere with the HMG box-DNA interaction (Fig. 6). Deletion of the N terminus of Lef-1 up to the position (amino acid 100) homologous to the start of the shorter TCF isoforms resulted in reduction of the transactivation capacity to a level comparable to that of these TCF-1 isoforms (Fig. 5C). Exchange of this region in human LEF-1 for the homologous region of TCF-1 further decreased the activity to a level comparable to the long TCF-1 isoforms (Fig. 5C). The shorter LEF-1 form and the TCF-1–LEF-1 chimera were as effective as short or long TCF-1 isoforms in abrogating transactivation mediated by wild-type LEF-1 (Fig. 5D). We concluded (i) that the N terminus of LEF-1 was crucial for its strong transactivating capacity in our system and that it could not be replaced by the homologous region of TCF-1; and (ii) that the TCF-1 N terminus did not interfere with DNA binding of the intact TCF-1 protein but that, both in the contexts of both TCF-1 and LEF-1, it blocked the transactivating potential of the N-terminally truncated proteins in the TCR-α enhancer assay.

DISCUSSION

The present study highlights the complexity of transcripts that emanate from the human and murine *TCF-1* genes. The presence of two promoters, of four alternative exons, and of three splice acceptors sites in the last exon yield a theoretical total of 96 different mRNAs. Direct sequencing has revealed the presence of all novel sequences on previously described genomic fragments encompassing the human *TCF-1* gene (23), underscoring the authenticity of the newly identified exons (Fig. 1).

The unexpected identification of an upstream promoter extends the known N terminus by 116 amino acids. This novel region is similar to Lef-1 (and chicken Tcf-1) throughout its length. The translation start site of the previously described short N-terminal forms of TCF-1 coincides with methionine residue 117 of the extended form of TCF-1. In addition, a newly identified exon (IVa) of TCF-1 is homologous to an alternative sequence found in human LEF-1 (29). C terminal to the HMG box, the homology to LEF-1 is completely lost. The LEF-1 protein actually carries a small constitutive C terminus. This contrasts with the multiple C termini of TCF-1. The TCF-1B and TCF-1E C termini (in that order) appear to be the most abundant sequences. TCF-1B is likely of functional importance because of the virtual identity to the C terminus of chicken Tcf (3). The conservation of TCF-1E C-terminal sequences between human and mouse *TCF-1* genes again implies a conservation of function.

The significance of the inclusion of exon IIa (with two in-frame stop codons) in *TCF-1* mRNA is currently unclear. Theoretically, the conceptual translation product is a 16-kDa peptide that could block normal functioning of the TCF-1 N terminus. As the function of this N terminus is different from that of the N terminus LEF-1 and remains to be resolved (see below), this notion can currently not be tested.

Western blotting of extracts prepared from sorted human thymocyte subpopulations has not revealed significant differences in abundance between individual TCF-1 isoforms (18). Thus, neither alternative splicing nor promoter activities appear to be regulated in a stage-specific fashion during T-lineage differentiation.

With the use of a reporter gene construct based on the T-cell-specific TCR-α enhancer, we show that a subset of TCF-1 isoforms transactivates transcription. Importantly, none of the isoforms can transactivate transcription through a concatemer of the recognition sequence (not shown). TCF-1 was 10-fold less potent in transactivating the TCR-α enhancer than LEF-1. However, the expression of TCF-1 mRNA in the thymus and in T-cell lines is up to 100-fold higher than that of LEF-1, as assessed by probing of cDNA libraries. Specifically, the frequency of TCF-1 clones ranges up to 1 per 200 to 500

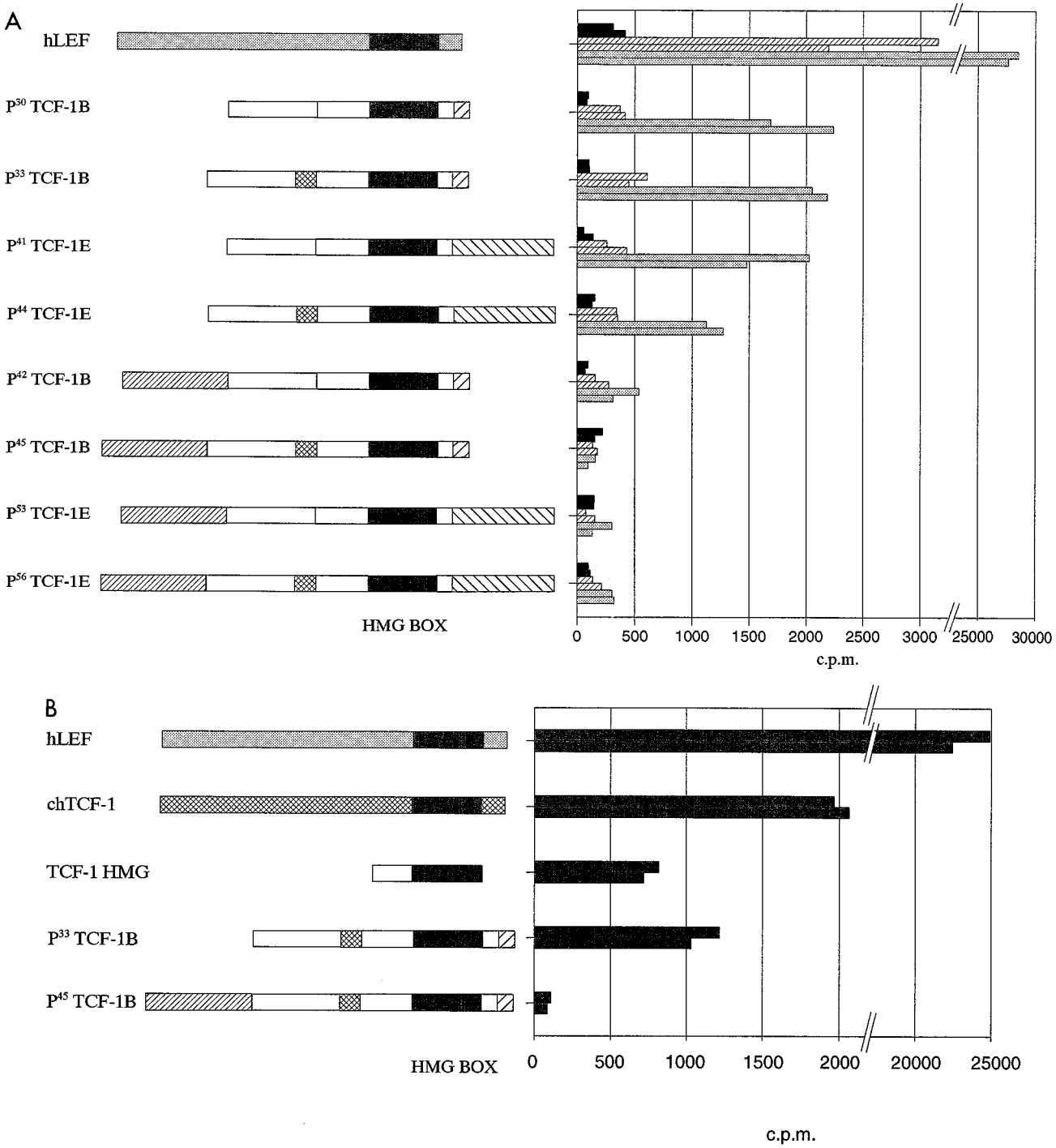

FIG. 5. TCF-1 protein isoforms display differential transactivating capacities. (A) Cotransfection of TCF-1 splice alternatives with a TCR-α enhancer–CAT construct (5, 12) in AZU2 B cells leads to an increase in transcription, exclusively observed upon transfection of the short N-terminal forms. Note the strict dependence of activity of the TCR-α enhancer in AZU2 B cells on the presence of transfected human LEF-1 (hLEF) or TCF-1. Transfections were performed in duplicate; the duplicate values are given as counts per minute of extractable [^{14}C]butyryl-chloramphenicol. Symbols for structures: ▨, extended N terminus, ▩, exon IVa; ▨, TCF-1B-specific C terminus; ▨, TCF-1E-specific C terminus. Symbols for bar graph: ■, 75 ng; ▨, 750 ng; ▨, 7,500 ng. (B) The HMG box of TCF-1 (residues 234 to 351) mediates a level of transactivation comparable to that observed with P^{33}TCF-1. Interestingly, chicken Tcf-1 (chTCF) also transactivates transcription through the TCR-α enhancer. Transfection of P^{45}TCF-1B serves as a negative control. Duplicate values are given. (C) AZU2 cells were transfected with the indicated expression constructs as described in Materials and Methods. As demonstrated, the truncated human LEF behaves like a short TCF-1 isoform in this assay, whereas the TCF-LEF chimera behaves like a long TCF-1 isoform. (D) AZU2 cells were cotransfected with 3 μg of LEF-1 expression vector in combination with 3 μg of control (pBluescript) or any of the indicated expression plasmids. The two artificial LEF-1 isoforms behave like the two TCF-1 isoforms with respect to a decrease in the measured LEF-1-mediated transactivation.

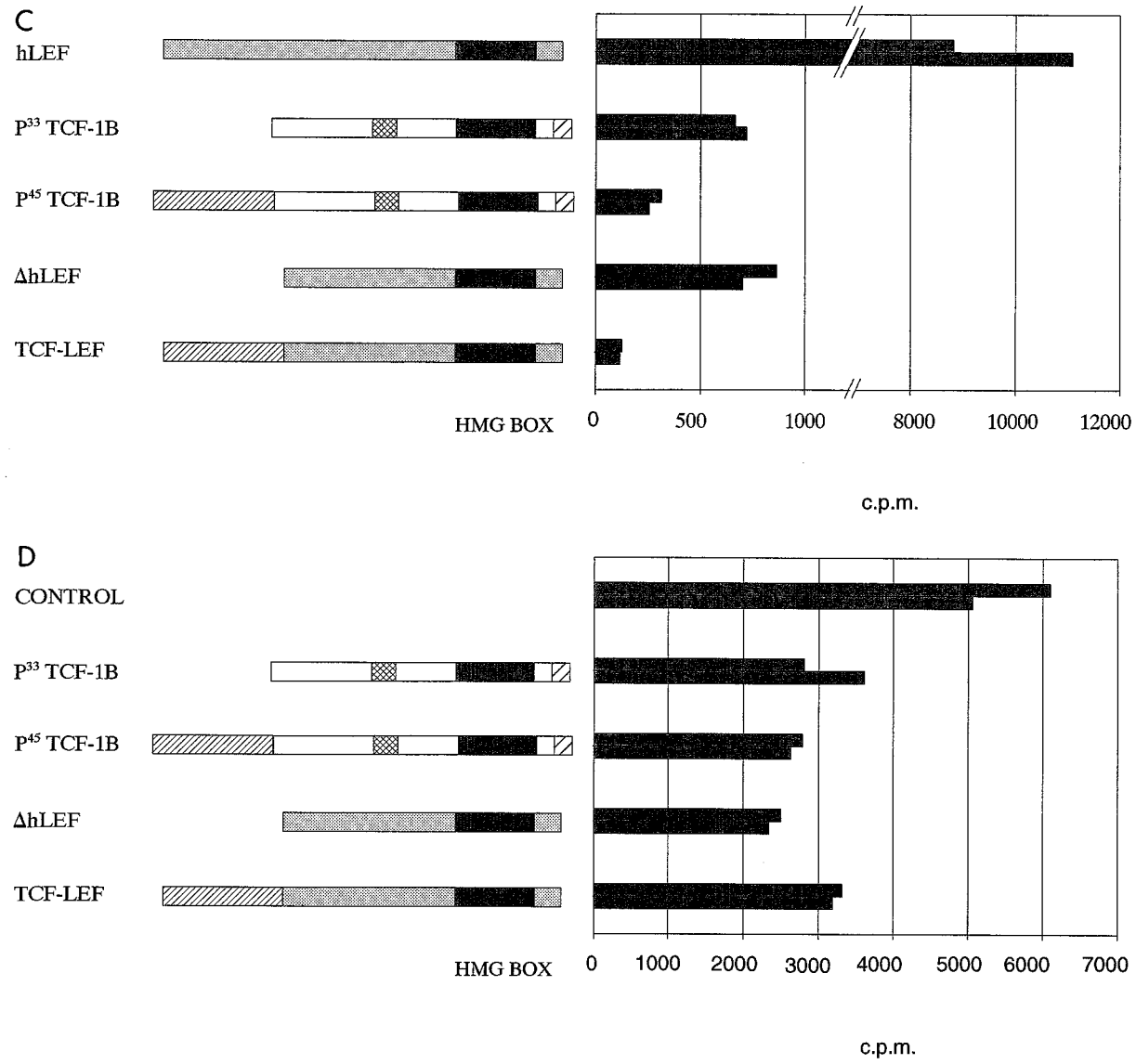

FIG. 5—*Continued.*

recombinant phage plaques, whereas LEF-1 frequencies in the same libraries were estimated at 1 per 25,000 plaques (21). We have also estimated TCF-1 protein levels in Jurkat cells to be 240 ng/10^6 cells. Waterman and Jones have reported a four-step purification of LEF-1 protein from Jurkat cells (30). Their yield was approximately 5 ng/10^6 cells. In the last two steps, a twofold decrease of LEF-1 protein occurred. On the basis of this report, a realistic estimate of the amount of LEF-1 protein would be 20 to 50 ng/10^6 cells. It appears reasonable to assume that the abundance of TCF-1 protein compensates for its poorer ability to transactivate the TCR-α enhancer compared with that of LEF-1. Given these considerations, we believe that TCF-1 and LEF-1 may perform redundant functions with respect to the control of TCR-α gene expression.

It is somewhat curious that the long TCF-1 isoforms do not transactivate transcription, despite the fact that they appear closer in sequence to LEF-1 than their smaller counterparts. We believe that the shorter forms of TCF-1 transactivate solely because of the interaction of the HMG box with the TCR-α enhancer. The additional effect of full-length LEF-1 derives mostly from its N terminus (references 1, 8, and 9 and this report). We demonstrate here that the N terminus of TCF-1 cannot replace that of LEF-1 in the TCR-α enhancer assay. The nonfunctional TCF-1 N terminus does not prohibit DNA binding. It does, however, interfere with context-specific transactivation on the TCR-α enhancer in transfected cells, possibly by steric hindrance. It appears likely that the functions of the N termini of LEF-1 and TCF-1 have diverged, the LEF-1 N terminus being relevant for the TCR-α enhancer and the TCF-1 N terminus being relevant for as yet unidentified target genes.

The functional redundancy between the shorter TCF-1 isoforms and LEF-1 provides an explanation for a paradox that has resulted from individual gene knockout experiments. *Tcf-1* knockmice display a severe impairment in thymocyte development, which is blocked at the immature single positive stage (28). Since the mutation is leaky, particularly around birth, peripheral T cells occur, albeit at reduced levels. These T cells

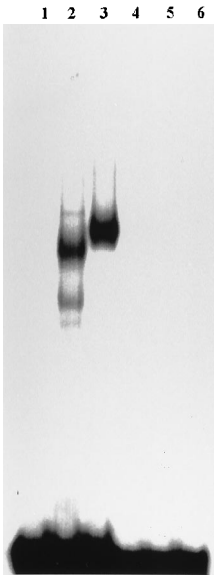

FIG. 6. The extended TCF-1 N terminus does not reconstitute an intramolecular DNA-binding inhibition domain. Gel retardation analysis of in vitro-translated TCF-1B isoforms was performed. Lanes 1 and 4, control lysate; lanes 2 and 5, P^{33}TCF-1B; lanes 3 and 6, P^{45}TCF-1B. Lanes 1 to 3 contain the TCF-1-binding oligonucleotide MW56; sequence specificity was confirmed by incubation with the nonbinding oligonucleotide MW56Sac (lanes 4 to 6). The amounts of TCF-1 protein per lane were comparable, as analyzed by Western blotting (not shown). The long and short TCF-1 proteins have similar binding affinities, as is evident from this experiment.

are normal with respect to phenotype and function. Specifically, TCR-α mRNA levels are unaffected. Along these lines, Van Genderen et al. recently created a *Lef-1*-deficient mouse strain (25). Disruption of the *Lef-1* gene affects hair follicle, tooth, mammary gland, and trigeminal nucleus development but, surprisingly, leaves the immune system intact (25). As in *Tcf-1*$^{-/-}$ mice, TCR-α expression in *Lef*-mutant mice is unaffected.

As described above, several studies have established that, at least when tested in transient transfections, TCR-α is strongly dependent on the TCF-LEF site and that LEF-1 is a potent, context-dependent transactivator for this enhancer (1, 8, 9). The current study predicts that TCR-α expression will be affected only in double-knockout mice. Such experiments are currently under way (14).

The exquisitely T-cell-specific phenotype of the *Tcf-1* knockout mice remains unresolved at the molecular level. The identification of numerous novel Tcf-1 isoforms now allows the creation of transgenic mice expressing single isoforms of Tcf-1 in order to evaluate which forms can complement the phenotype. Molecular analysis of wild-type, mutant, and mutant-complemented mice will allow the identification of crucial genes, downstream of Tcf-1 in the T-lineage differentiation pathway, that are dependent on proper Tcf-1 function.

ACKNOWLEDGMENTS

We thank M. van Dijk, M. Schilham, and J. van de Winkel for helpful comments on the manuscript and R. Grosschedl for making available the TCR-α enhancer CAT construct.

This study was financially supported by a PIONIER grant from NWO to H.C.

REFERENCES

1. **Carlsson, P., M. L. Waterman, and K. A. Jones.** 1993. The hLEF/TCF-1α HMG protein contains a context-dependent transcriptional activation domain that induces the TCRα enhancer in T cells. Genes Dev. **7:**2418–2430.
2. **Castrop, J.** Unpublished data.
3. **Castrop, J., R. Hoevenagel, J. R. Young, and H. Clevers.** 1992. A common ancestor of the mammalian transcription factors TCF-1 and TCF-1α/LEF-1 expressed in chicken T cells. Eur. J. Immunol. **22:**327–1330.
4. **Castrop, J., D. van Wichen, M. van de Wetering, M. Koomans-Bitter, R. de Weger, J. van Dongen, and H. Clevers.** The human TCF-1 gene encodes a nuclear DNA-binding protein, uniquely expressed in normal and neoplastic T lineage lymphocytes. Blood, in press.
5. **Clevers, H. C., M. A. Oosterwegel, and K. Georgopoulos.** 1993. Transcription factors in early T cell development. Immunol. Today **14:**591–596.
6. **Dooijes, D., M. van de Wetering, L. Knippels, and H. Clevers.** 1993. The *Schizosaccharomyces pombe* mating-type gene *mat*-MC encodes a sequence-specific DNA-binding HMG box protein. J. Biol. Chem. **268:**24813–24817.
7. **Giese, K., J. Cox, and R. Grosschedl.** 1992. The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein complexes. Cell **69:**185–195.
8. **Giese, K., and R. Grosschedl.** 1993. LEF-1 contains an activation domain that stimulates transcription only in a specific context of factor-binding sites. EMBO J. **12:**4667–4676.
9. **Giese, K., C. Kingsley, J. R. Kirshner, and R. Grosschedl.** 1995. Assembly and function of a TCRα enhancer complex is dependent on LEF-1-induced DNA bending and multiple protein-protein interactions. Genes Dev. **9:**995–1008.
10. **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 (London) **344:**830–836.

10a. **Korinek, ●., and H. Clevers.** Unpublished data.

11. **Laudet, V., D. Stehelin, and H. C. Clevers.** 1993. Ancestry and diversity of the HMG box superfamily. Nucleic Acids Res. **21:**2493–2501.
12. **Leiden, J.** 1993. Transcriptional regulation of T cell receptor genes. Annu. Rev. Immunol. **11:**539–588.
13. **Love, J. J., X. Li, D. A. Case, K. Giese, R. Grosschedl, and P. E. Wright.** 1995. Structural basis for DNA bending by the architectural transcription factor LEF-1. Nature (London) **376:**791–795.
14. **Okamura, R. M., S. Verbeek, R. Grosschedl, and H. Clevers.** Unpublished data.
15. **Oosterwegel, M., M. van de Wetering, D. Dooijes, L. Klomp, A. Winoto, K. Georgopoulos, F. Meijlink, and H. Clevers.** 1991. Cloning of murine TCF-1, a T cell-specific transcription factor interacting with functional motifs in the CD3-ε and T cell receptor α enhancers. J. Exp. Med. **173:**1133–1142.
16. **Oosterwegel, M., M. van de Wetering, J. Timmerman, A. Kruisbeek, O. Destree, F. Meijlink, and H. Clevers.** 1993. Differential expression of the HMG box factors TCF-1 and LEF-1. Development **118:**439–448.
17. **Sheridan, P. L., C. T. Sheline, K. Cannon, M. L. Voz, M. J. Pazin, J. T. Kadonaga, and K. A. Jones.** 1995. Activation of the HIV-1 enhancer by the LEF-1 HMG protein on nucleosome assembled DNA in vitro. Genes Dev. **9:** 2090–2104.
18. **Staal, F., H. Spits, T. Castrop, and H. Clevers.** Unpublished data.
19. **Travis, A., A. Amsterdam, C. Belanger, and R. Grosschedl.** 1991. Lef-1, a gene encoding a lymphoid-specific protein, with an HMG domain, regulates T-cell receptor α enhancer function. Genes Dev. **5:**880–894.
20. **van de Wetering, M.** Unpublished data.
21. **van de Wetering, M. Oosterwegel, and H. Clevers.** Unpublished data.
22. **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 J. **10:**123–132.
23. **van de Wetering, M., M. Oosterwegel, F. Holstege, D. Dooijes, R. Suikerbuik, A. Geurts van Kessel, and H. Clevers.** 1992. The human T cell transcription factor-1 gene. J. Biol. Chem. **267:**8530–8536.
24. **van de Wetering, M., M. Oosterwegel, K. van Norren, and H. Clevers.** 1993. Sox-4, an Sry-like HMG box protein, is a transcriptional activator in lymphocytes. EMBO J. **12:**3847–3854.
25. **Van Genderen, C., R. M. Okamura, I. Farinas, R. G. Quo, T. G. Parslow, L. Bruhn, and R. Grosschedl.** 1994. Development of several organs that require inductive epithelial-mesenchymal interactions is impaired in LEF-1 deficient mice. Genes Dev. **8:**2691–2704.
26. **van Houte, L. P. A., V. P. Chuprina, M. van de Wetering, R. Boelens, R. Kaptein, and H. C. Clevers.** Solution structure of the sequence-specific HMG box of the lymphocyte transcriptional activator Sox-4. J. Biol. Chem. in press.
27. **Verbeek, S., and H. Clevers.** Unpublished data.
28. **Verbeek, S., D. Izon, F. Hofhuis, E. Robanus-Maandag, H. te Riele, M. van de Wetering, M. Oosterwegel, A. Wilson, H. R. MacDonald, and H. Clevers.** 1995. An HMG-box-containing T-cell factor required for thymocyte differentiation. Nature (London) **374:**70–74.
29. **Waterman, M. L., W. H. Fischer, and K. A. Jones.** 1991. A thymus-specific member of the HMG protein family regulates the human T cell receptor Cα enhancer. Genes Dev. **5:**656–669.
30. **Waterman, M. L., and K. 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:**621–636.
31. **Werner, M. H., J. R. Huth, A. M. Gronenborn, and G. M. Clore.** 1995. Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. Cell **81:** 705–714.