A common ancestor of the mammalian transcription factors TCF-1 and TCF-1α/LEF-1 expressed in chicken T cells*

Several mammalian T cell-specific transcription factors have been cloned recently. Two of these, TCF-1 and TCF-1α/LEF-1, display a moderate level of overall sequence similarity, and contain virtually identical versions of a novel type of DNA-binding domain, the HMG box. To study evolutionary aspects of the TCF transcription factors in relation to lymphoid differentiation, we have isolated chicken TCF clones from a spleen cDNA library. Low-stringency screening with human probes as well as a polymerase chain reaction-aided strategy resulted in the cloning of a single chicken TCF (chTCF) gene. Sequence comparison revealed that chTCF contained a TCF-1α-like N terminus, and a TCF-1α-like C terminus. Furthermore, TCF-1 and TCF-1α were more homologous to chTCF than to each other. We postulate that chTCF is the direct descendant of a single ancestral gene, which has been duplicated in mammals to yield TCF-1 and TCF-1α.

1 Introduction

Tissue-specific transcription factors are generally assumed to play a central role in differentiation [1–3]. The availability of well-defined cell lines, antibodies and cloned genes renders the lymphocyte lineage an attractive model system for the study of transcription control during vertebrate development. The characterization of several T lymphocyte-specific enhancers has allowed the cloning of transcription factors controlling T cell-specific gene expression [4–8]. Two such factors, TCF-1 [4, 5] and TCF-1α/LEF-1 [6, 7], are members of a novel gene family, characterized by a region of approximately 80 amino acids with homology to high mobility group I (HMG) proteins. This domain, termed the HMG box, mediates sequence-specific DNA binding [5, 9]. This family further comprises the mammalian sex-determining gene SRY [10], the polymerase I transcription factor UBF [9], and mating-type genes of Schizosaccharomyces pombe and Neurospora crassa [11, 12].

TCF-1 and TCF-1α contain very similar HMG boxes (98 % amino acid identity), and presumably bind identical sequence motifs. TCF-1 expression appears to be limited to the T lymphocyte lineage [5], whereas TCF-1α/LEF-1 is also expressed by pre-B cells [7]. TCF-1 was originally cloned based on its specificity for the CD3-e enhancer [4] and was subsequently demonstrated to bind to TCR-α, -β and -ε enhancers [5, 13]. TCF-1α/LEF-1 binds to a number of motifs including that in the TCR-α enhancer [6, 7]. Based on methylation interference footprinting and on a comparison of cognate sites, the TCF-1 HMG box is proposed to recognize the degenerate motif A/T A/T/CANAG [4, 13]. To gain insight in the evolution of the lymphoid-specific TCF genes, we have pursued the identification of homologues in birds.

2 Methods

2.1 cDNA library

Spleens were obtained from 10-week-old White Leghorn RPL Line 7 chickens (Regional Poultry Laboratory, East Lansing, MI). Spleen cells were depleted of red blood cells by centrifugation on Ficoll. Adherent cells and B lymphocytes were removed by adherence to nylon wool. Poly-A RNA was prepared from total RNA as described previously [14]. cDNA was synthesized using a BRL cDNA kit (BRL, Oxbridge, GB) and oligo-dT primers; BstXI-linkers were added [14, 15]. The cDNA was then size-fractionated over agarose and cloned into pCDMS [15]. The library was transformed into E.coli MC1061/p3 by electroporation; five independent sublibraries of approximately 2 × 10^6 colonies each were obtained.

2.2 Polymerase chain reaction (PCR)-aided cloning of chTCF

Two degenerate primers were designed based on the extensive homology between the HMG boxes of TCF-1 and TCF-1α. The sequences of the “guesser”-primers were: GGAATTCAATGC(GATC)TT(TC)ATG(TC)GAGGTCGGAATG (positive strand); and GGAAGCTTACG(GA)TAATG(TA)(AT)CG(TATC)GAGGATTCCG (negative strand), where bases at degenerate positions are indicated between parentheses. The relative position of the primers in the HMG box of chTCF is indicated in Fig. 1. PCR reactions were performed on plasmid DNA derived from the five sublibraries. Conditions: 10 ng of plasmid DNA was amplified for 25 cycles

* This work was supported by the Dutch Foundation for Research NWO.

Correspondence: Hans C. Clevers, Department of Immunology, University Hospital, P.O.Box 85500, NL-3508 GA, Utrecht, The Netherlands

© VCH Verlagsgesellschaft mbH, D-6940 Weinheim, 1992

0014-2980/92/0505-1327$3.50 + .25/0

I 10146

[1] 10146
using 250 ng of each primer. Annealing: 1 min at 42°C; extension: 2 min at 65°C; melting: 1 min at 95°C. All reagents (including Taq polymerase) were from Promega (Madison, WI). PCR products were digested with Hind III and Eco RI, cloned into pBluescript KS (Stratagene, San Diego, CA) and sequenced.

2.3 Low-stringency screening

Sublibraries were plated out (10⁶ colonies per sublibrary), lifted in duplicate onto nitrocellulose, and probed with human TCF-1 (the Eco RI insert of TCF-1 [4]), and with a 613-bp internal Sal I fragment of human TCF-1α [6] and M. Oosterwegel and H. C.C., unpublished) at 35°C in 50% formamide-containing prehybridization mixture [4]. Filters were washed extensively with 4 × SSC, 1% SDS, at 50°C. Colonies positive in duplicate were picked and rendered clonal by secondary screening.

2.4 Northern blotting

Total RNA was extracted from freshly isolated chicken tissues using RNazol (Cinna-Biotech, Houston, TX) and 5 mg was subjected to Northern analysis as described elsewhere [5]. As a probe, the insert of the longest chTCF cDNA clone was used. The sequence of this clone, pchTCF is represented in Fig. 1.

3 Results and discussion

Sequences encoding TCF-like genes from the human, mouse, and Drosophila genome (J. Castrop, unpublished). Ten chicken PCR clones (two per sublibrary) were analyzed and were found to be of identical sequence (not shown). Next, the spleen cDNA library was probed with a mixture of two probes representing the HMG boxes of human TCF-1 and TCF-1α. Nine independent clones were isolated. Sequence analysis of the clones revealed that they were all derived from the same gene, which was termed chTCF. The sequence of the HMG box of chTCF was identical to that of the previously obtained PCR clones. The chTCF gene encoded a 393-amino acid polypeptide, containing an HMG box with >96% similarity to mammalian TCF-1 and TCF-1α (Figs. 1 and 2). The overall structure of chTCF was very similar to that of TCF-1 and TCF-1α: the N-terminal two thirds of the predicted protein is rich in prolines, whereas the C-terminal third was predominantly basic.

Comparison of chTCF with mammalian TCF-1 and TCF-1α (Fig. 2) revealed several noteworthy features: (a) the N terminus of chTCF was similar to that of TCF-1α; (b) Met-96 of chTCF coincided with Met-1 in TCF-1. C-terminal to this residue, the three proteins aligned well with a particularly high level of identity in the HMG box region; (c) the 25-amino acid C-terminal region of chTCF corresponded to the C terminus of TCF-1. The TCF-1α C terminus appeared unrelated; (d) in the common stretch N-terminal to the HMG box (96 to 277 in chTCF), identity between chTCF and TCF-1α occurred at 110 positions, and between chTCF and TCF-1 at 104 positions. In contrast, TCF-1α and TCF-1 only shared 71 identical residues in this region. Areas of identity of chTCF to either TCF-1α or TCF-1 were distributed patch-wise over the region. Thus, chTCF appeared to be a homologue of the mammalian TCF genes, with a TCF-1α-like N terminus and a TCF-1-like C terminus. Importantly, TCF-1 and TCF-1α were more similar to chTCF than to each other.
Northern blot analysis (Fig. 3) on RNA obtained from a number of chicken tissues demonstrated that chTCF was only detectably expressed in the thymus. Notably, the bursa did not contain chTCF mRNA. Thus, the expression of chTCF coincided with the pattern of expression that we documented earlier for mouse TCF-1 [5]. The absence of expression in B lineage cells in the bursa contrasts with that reported for murine TCF-1α [7], but is in line with our observations for human and murine TCF-1 (M. Oosterwegel and H.C.C., unpublished). The conservation of the T lymphoid expression pattern in mammals and birds stresses the functional importance of the TCF genes for this cell type.

Our efforts have failed, yet to yield evidence for the expression of a second TCF-like gene in chicken T cells. Unfortunately, analysis at genomic level is hampered by the presence of at least two additional TCF-like genes in the vertebrate genome (J.C. and H.C.C., submitted). These TCF-3 and -4 genes are not expressed in lymphoid cells, and, although more distantly related, display >92% similarity towards TCF-1 and TCF-1α in the HMG box region.

Taken together, our data imply that chicken T cells (either predominantly or exclusively) express a TCF homologue with sequence characteristics of a putative ancestor of the mammalian lymphoid TCF genes. Duplication of this ancestral gene would have occurred after the divergence of birds and mammals. In theory, chTCF could also have originated from an unequal crossing-over between TCF-1 and TCF-1α. However, if that were the case, sequence comparison should be expected to reveal a pattern of high similarity to TCF-1α N-terminal to the hypothesized crossover point, and to TCF-1 C-terminal to that point. The patch-wise, alternating pattern of homology observed in this study, is at variance with such a mechanism of gene deletion.
Many transcription factors described to date belong to gene families with members sharing structurally and functionally highly similar DNA-binding domains [3]. It remains enigmatic, how such redundancy results in the tight control of tissue- and stage-specific transcription in metazoan organisms. The present study provides an example of a recent duplication event within a transcription factor gene family. The resulting two genes subsequently diverged with respect to sequence of the encoded proteins, which likely correlates with divergence of function. Strikingly, the DNA-binding HMG box remained essentially identical. The comparative study of the "lymphoid" TCF genes in birds and mammals might, thus, provide insight into the mechanisms that allow duplication of transcription factor genes with simultaneous conservation of the encoded DNA binding domains and diversification of function.

The authors wish to thank Dr. P. McLaughlan for providing chicken splenocytes, Dr. R. van Tilburg for chicken tissue samples, and Drs. J. van de Winkel, T. Logtenberg and S. Verbeek for critically reading the manuscript.

Received November 19, 1991; in revised form January 24, 1992.

4 References


