The EMBO Journal vol.12 no.10 pp.3847–3854, 1993

Sox-4, an Sry-like HMG box protein, is a transcriptional activator in lymphocytes

Marc van de Wetering, Mariëtte Oosterwegel, Klaske van Norren and Hans Clevers[1]

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

[1]Corresponding author

Communicated by N.Hastie

Previous studies in lymphocytes have described two DNA-binding HMG box proteins, TCF-1 and LEF-1, with affinity for the $^{A}/_{T}{}^{A}/_{T}$CAAAG motif found in several T cell-specific enhancers. Evaluation of co-transfection experiments in non-T cells and the observed inactivity of an AACAAAG concatamer in the TCF-1/LEF-1-expressing T cell line BW5147, led us to conclude that these two proteins did not mediate the observed enhancer effect. We therefore searched for additional HMG box proteins. By a PCR-aided strategy, we cloned *Sox-4*, a gene with homology to the HMG box region of the sex determining gene *SRY*. Sox-4 was expressed in T and pre-B lymphocyte lines and in the murine thymus. Significantly, BW5147 T cells did not express Sox-4. Recombinant Sox-4 bound with high affinity (K_d 3×10^{-11} M) to the minor groove of the AACAAAG motif, most likely contacting all seven base pairs. In contrast with observations on *TCF-1* and *LEF-1*, co-transfection with *Sox-4* unveiled a transactivating capacity, which mapped to its serine-rich C terminus. This region remained functional upon grafting onto a GAL4 DNA-binding domain. Sox-4 is thus the first 'classical' transcription factor in the *Sox* gene family with separable DNA-binding and transactivation domains. Our observations indicate that a detailed understanding of T cell-specific gene control must integrate the concerted activity of at least three tissue-specific HMG box genes.

Key words: HMG box/*Sox-4*/*SRY*/T cell differentiation/transactivation

Introduction

The high mobility group-1 (HMG) box was originally identified by Tjian and co-workers in the transcription factor UBF as a region of homology to HMG-1 proteins (Jantzen *et al.*, 1990). UBF reportedly contained four such regions of ~80 amino acids; one of these boxes was shown to mediate DNA binding. More than 60 HMG box proteins have since been reported and/or entered into sequence data bases (Laudet *et al.*, 1993). The HMG box proteins can be broadly divided into a group that recognizes DNA in a relatively non-specific fashion and a group that displays high sequence specificity. The former group includes amongst others UBF, HMG-1 and MT-TF1 (Wen *et al.*, 1989; Jantzen *et al.*, 1990; Parisi and Clayton, 1991). The latter group contains the products of several fungal genes implied in mating-type determination (Kelly *et al.*, 1988; Staben and Yanofsky, 1990; Sugimoto *et al.*, 1991), the mammalian sex determining gene *SRY* and related products (Gubbay *et al.*, 1990; Sinclair *et al.*, 1990), and two putative regulators of lymphoid differentiation TCF-1 (Oosterwegel *et al.*, 1991; van de Wetering *et al.*, 1991) and LEF-1 (Travis *et al.*, 1991; Waterman *et al.*, 1991). Computer-aided construction of an evolutionary tree of HMG box proteins has corroborated the dichotomy into non-specific *vis-à-vis* sequence-specific HMG boxes (Laudet *et al.*, 1993). Interestingly, the sequence-specific HMG boxes characterized to date display high affinity to the $^{A}/_{T}{}^{A}/_{T}$CAAAG motif despite a low level of amino acid homology (typically <25% identity). Sequence specificity has been best described for the *Schizosaccharomyces pombe* transcription factor Ste11 (Sugimoto *et al.*, 1991), *SRY* and a related rat gene *IRE-ABP* (Nasrin *et al.*, 1991; Harley *et al.*, 1992), the lymphoid proteins TCF-1 (Oosterwegel *et al.*, 1991a,b; van de Wetering *et al.*, 1991) and LEF-1 (Travis *et al.*, 1991; Waterman *et al.*, 1991), and the *S.pombe* mating type gene *MatMc* (Dooijes *et al.*, 1993). As demonstrated for LEF-1, TCF-1 and *SRY* (Ferrari *et al.*, 1992; Giese *et al.*, 1992; van de Wetering and Clevers, 1992), the sequence-specific binding of the HMG box occurs in an unusual fashion, i.e. in the minor groove, and induces a strong bend into the DNA helix.

Both TCF-1 and LEF-1 were identified in a search for transcription factors controlling lymphoid differentiation. *TCF-1* was originally cloned based on its affinity for the AACAAAG motif in the enhancer of the gene encoding CD3-ϵ, one of the constituent chains of the T cell receptor–CD3 complex (van de Wetering *et al.*, 1991). Human LEF-1 was originally identified as a T cell-specific protein binding to the TTCAAAG motif in the TCR-α enhancer (Waterman *et al.*, 1991). Murine *LEF-1* was cloned from pre-B lymphocytes by a subtraction strategy, and was shown to bind to the same motif (Travis *et al.*, 1991). Functional TCF-1/LEF-1 motifs have since been found in several other lymphoid-specific enhancers, including those of the T cell receptor (TCR)-β and TCR-δ genes (Oosterwegel *et al.*, 1991b) and of the CD4 gene (Sawada and Littman, 1991).

We have since demonstrated that within the haemopoietic system *TCF-1* is uniquely expressed by all T lineage cells beyond the prothymocyte stage, whereas *LEF-1* is expressed by all T lineage cells as well as by pro- and pre-B cells. Additionally, we detected a restricted and partially overlapping expression pattern of the two genes in several other organs during embryogenesis (Oosterwegel *et al.*, 1993). The two genes appear to derive from a recent gene duplication event as evidenced by the cloning of a single chicken homologue, chTCF (Castrop *et al.*, 1992a).

Our original interest in TCF-1 was based both on its T cell-specific expression and on the observation that a concatamer of its cognate motif acted as a 5- to 10-fold T cell-specific enhancer (van de Wetering *et al.*, 1991). Surprisingly, co-transfection of *TCF-1* and/or *LEF-1* with a reporter gene under control of the concatamer into non-T cells did not result in any significant transactivation (M. Oosterwegel and H.Clevers, unpublished). Even in the highly efficient COS cell expression system, our originally reported 2- to 3-fold transactivation effect (van de Wetering *et al.*, 1991) has since proved inconsistent. Similarly, co-transfection of *LEF-1* with its TCR-α minimal binding motif (TTCAAAG) concatamerized in a CAT–reporter construct did not result in transactivation (Travis *et al.*, 1991). Moreover, we found that the AACAAAG concatamer was inactive in the murine T lineage cell line BW5147 (this study), despite the observation that this cell line expresses both TCF-1 and LEF-1.

At least two hypotheses could explain these discrepancies. (i) Transactivation of the AACAAAG motif involves an as yet unidentified transcription factor. This putative factor is probably another T cell-specific HMG box protein, absent or mutated in BW5147. (ii) TCF-1 and/or LEF-1 co-operate with an adaptor protein, itself uniquely present in T cells, but absent in BW5147. To test the first hypothesis, we have adopted a PCR-based strategy for cloning of novel HMG box genes based on homology to known sequence-specific HMG boxes. Our original attempts aimed at isolating *TCF-1/LEF-1*-like genes yielded two highly homologous human genes, *TCF-3* and *-4*. However, these genes were not expressed in lymphoid cells (Castrop *et al.*, 1992b). The present study pursues the identification of *SRY*-like ('*Sox*') genes expressed in T lymphocytes, responsible for transcriptional transactivation through the AACAAAG motif.

Results

Identification of Sox sequences expressed in murine T cells

Our interest in the AACAAAG motif derived from the original observation that a concatamer of this minimal motif functioned as a T cell-specific enhancer (van de Wetering *et al.*, 1991). Since the two known lymphoid-specific AACAAAG-binding proteins TCF-1 and LEF-1 did not transactivate transcription through this concatamer, we searched for a third gene encoding a protein with similar DNA-binding characteristics.

It has recently been demonstrated that the product of the mammalian sex-determining gene *SRY* binds with high affinity to the minor groove of the AACAAAG motif (Harley *et al.*, 1992; van de Wetering and Clevers, 1992). Numerous *Sox* genes have been cloned based on sequence homology to the SRY HMG box (Gubbay *et al.*, 1990; Denny *et al.*, 1992a; Wright *et al.*, 1993). At least two of these, mouse *Sox-5* and rat *IRE-ABP*, reportedly bind to the AACAAAG motif (Nasrin *et al.*, 1991; Denny *et al.*, 1992). Since affinity for the AACAAAG motif probably represents a general characteristic of Sry-like HMG boxes, we attempted cloning of *Sox* genes from cDNA derived from a mature mouse T cell line by a PCR-based strategy. Guess-mer primers were designed based on the homology between the HMG boxes of Sry and Sox-1 to -4 (Gubbay *et al.*, 1990). To allow very low stringency PCR amplification, we contrived a 'reverse' blue-white screening method after subcloning PCR products between the *Hind*III and *Eco*RI sites of pBluescriptSK. Insertion of a cDNA encoding an HMG box would result in maintenance of the *lacZ* reading frame and thus in the growth of blue colonies. Insertion of random amplified sequences would generally result in white colonies due to reading frame shifts or stop codons. This approach proved very powerful. Only 10% of the colonies were blue; subsequent analysis of these colonies showed that >95% of these (recombinant) colonies potentially encoded HMG boxes. None of the 10 white colonies analysed contained HMG box sequences.

A

```
     agagcagcagagtgaggggaagagggccgtctccctcccggtttccagttcttgcacgct   60
     gtttcttagagagtctgcagtgggggaactctgccggtaaccagctccccttcttgcagg   120
     agggagggagaaacatacatttattcatgccggtctgttgcatgcaagcttcttggcttc   180
     ctaccttgcaacaaaataattgcaccaactcctcagcgccgattccgcccacagagagtc   240
                                                         *
     ccggagccagagtcgctttggctttgcactgcaggaaagggacttaggcgctagagacga   300
     tgtcgctttcctgagctaccgcgagctctcgtgaactgcaatcgactgcttcagggaaag   360
     gggtgggggaaagacttgccccggaggcggcgagaaacttgcgtttggaagatactccgg   420
     ctaccaacgtttggagaaactcctcgccgcggccgactccagcctcggtgcccgcaggag   480
     cacttcagcggtgagggaggacggagggcctcggggactctaggttggcggcgggaggcg   540
     gccgcccctggcccgcgcgccgctcaggggaccctggtctccgcccccggggagcccgcag   600
     gcccggcgaccgcgccgcgagcgtgtgagcgcgcgtgggcgccggcaagccggggccATG   660
  1                                                            M
     GTACAACAGACCAACAACGCGGAGAACACTGAGGCTCTGCTGGCCGGGGAGAGCTCGGAC   720
  2   V  Q  Q  T  N  N  A  E  N  T  E  A  L  L  A  G  E  S  S  D
     TCGGGCGCCGGCCTGGAGCTGGGCATCGCGTCCTCCCCGACGCCTGGCTCCACCGCGTCG   780
 22   S  G  A  G  L  E  L  G  I  A  S  S  P  T  P  G  S  T  A  S
     ACGGGCGGCAAGGCGGACGACCCCAGCTGGTGCAAGACGCCCAGTGGCCACATCAAGCGG   840
 42   T  G  G  K  A  D  D  P  S  W  C  K  T  P  S  G  H  I  K  R
     CCCATGAACGCCTTTATGGTGTGGTCGCAGATCGAGCGGCGCAAGATCATGGAGCAGTCG   900
 62   P  M  N  A  F  M  V  W  S  Q  I  E  R  R  K  I  M  E  Q  S
     CCCGACATGCACAACGCCGAGATCTCCAAGCGGCTAGGCAAACGCTGGAAGCTGCTCAAG   960
 82   P  D  M  H  N  A  E  I  S  K  R  L  G  K  R  W  K  L  L  K
     GACAGCGACAAGATTCCGTTCATCCAGGAGGCGGAGCGCCTGCGCCTCAAGCACATGGCT   1020
102   D  S  D  K  I  P  F  I  Q  E  A  E  R  L  R  L  K  H  M  A
     GACTACCCTGACTACAAGTACCGGCCGCGAAAGAAGGTGAAGTCGGGCAACGCGGGCGCG   1080
122   D  Y  P  D  Y  K  Y  R  P  R  K  K  V  K  S  G  N  A  G  A
     GGATCGGCGGCCACAGCCAAGCCAGGGGAGAAGGGCGACAAGGTCCGGGGCAGCAGCGGC   1140
142   G  S  A  A  T  A  K  P  G  E  K  G  D  K  V  R  G  S  S  G
     CACGCGGGAAGCAGCCACGCGGGGGGTGGCGCGGGCGGCAGCTCCAAGCCCGCGCCCAAG   1200
162   H  A  G  S  S  H  A  G  G  G  A  G  G  S  S  K  P  A  P  K
     AAGAGCTGTGGCCCCAAGGTGGCGGGCAGCTCGGTCGGCAAGCCCCACGCTAAGCTGGTC   1260
182   K  S  C  G  P  K  V  A  G  S  S  V  G  K  P  H  A  K  L  V
     CCGGCGGGCGGCAGCAAGGCGGCTGCATCGTTCTCTCCAGAGCAAGCTGCCCTGCTGCCC   1320
202   P  A  G  G  S  K  A  A  A  S  F  S  P  E  Q  A  A  L  L  P
     CTGGGGGAGCCCACGGCCGTCTACAAGGTGCGGACTCCCAGTCGGGCCACTCCGGCCGCC   1380
222   L  G  E  P  T  A  V  Y  K  V  R  T  P  S  R  A  T  P  A  A
     TCCTCCTCGCCGTCCAGTGCGCTGGCCACCCCAGCCAAACACCCTGCCGACAAGAAAGTG   1440
242   S  S  S  P  S  S  A  L  A  T  P  A  K  H  P  A  D  K  K  V
     AAGCACGTCTACCTGTTTGGAAGCCTGGGCGCTTCGGCGTCTCCCGTCGGGGGCCTGGGT   1500
262   K  H  V  Y  L  F  G  S  L  G  A  S  A  S  P  V  G  G  L  G
     GCGTGCGCCGACCCCAGTGATCCACTGGGGTTGTACGAAGATGGAGGCCCGGGATGCTCG   1560
282   A  C  A  D  P  S  D  P  L  G  L  Y  E  D  G  G  P  G  C  S
     CCCGATGGCCGGAGTCTGAGCGGCCGCAGCAGCGCAGCATCATCGCCAGCCGCCAGCCGA   1620
302   P  D  G  R  S  L  S  G  R  S  S  A  A  S  S  P  A  A  S  R
     TCGCCCGCTGACCACCGCGGCTACGCCAGCCTACGCGCAGCCTCGCCCGCCCCGTCCAGC   1680
322   S  P  A  D  H  R  G  Y  A  S  L  R  A  A  S  P  A  P  S  S
     GCGCCCTCGCACGCGTCCTCCTCGCTCTCCTCGTCCTCTTCCTCCTCCTCGGGCTCTTCG   1740
342   A  P  S  H  A  S  S  S  L  S  S  S  S  S  S  S  S  G  S  S
     TCGTCCGACGACGAGTTCGAAGACGACCTGCTCGACCTGAACCCCAGCTCAAACTTTGAG   1800
362   S  S  D  D  E  F  E  D  D  L  L  D  L  N  P  S  S  N  F  E
     AGCATGTCCCTGGGCAGTTTCAGCTCCTCATCGGCGCTCGATCGGGACCTGGATTTTAAC   1860
382   S  M  S  L  G  S  F  S  S  S  S  A  L  D  R  D  L  D  F  N
     TTCGAACCCGGCTCAGGCTCCCACTTCGAATTCCCGGACTATTGCACGCCCGAGGTGAGC   1920
402   F  E  P  G  S  G  S  H  F  E  F  P  D  Y  C  T  P  E  V  S
     GAGATGATCTCGGGAGATTGGCTGGAGTCCAGCATCTCTAACCTGGTCTTCACCTACtga   1980
422   E  M  I  S  G  D  W  L  E  S  S  I  S  N  L  V  F  T  Y
     agggagcgcgggccggggagaaggtgggccaagaggcaggagaggagagagggaaaaaag   2040
     aaacaaaaaacaaacaaaaaaaaaaaaaaga                                2100
```

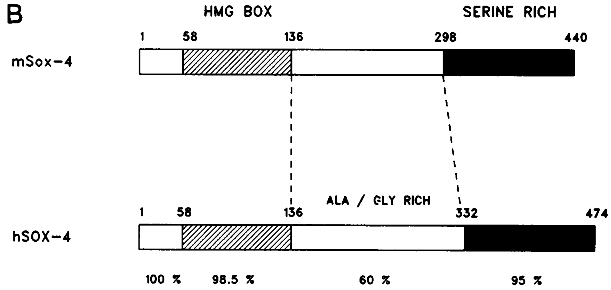
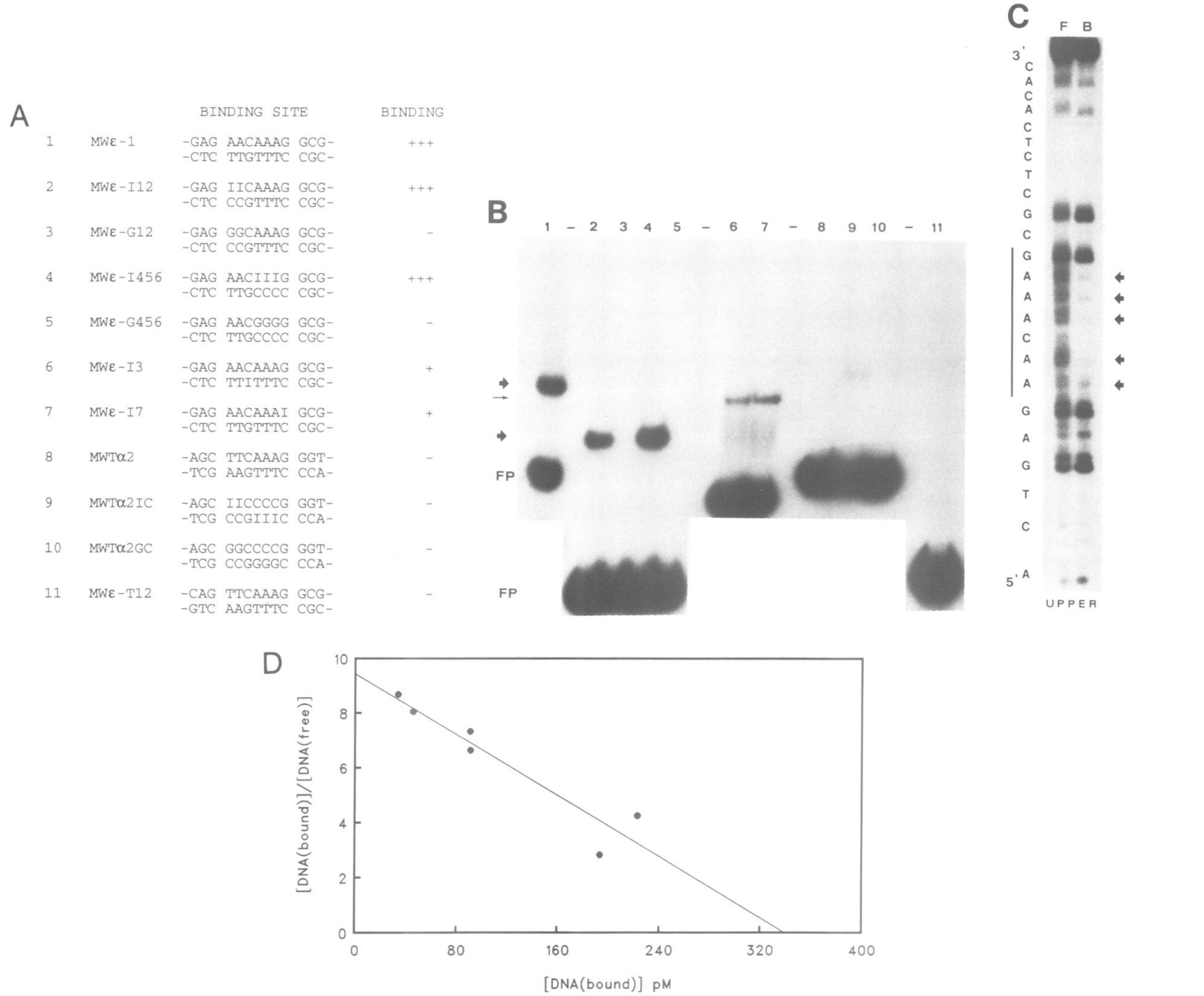


Fig. 1. Isolation of *Sox-4* cDNA clones. (**A**) Sequence of murine Sox-4 (EMBL accession number X70298). The HMG box is underlined. An in-frame stop codon is indicated with an asterisk. Bases are numbered on the right; amino acids on the left. (**B**) Graphic representation of the murine and human Sox-4 proteins. Percentages of amino acid identity are given below hSox-4.

The sequences of 70 randomly picked blue colonies were determined. More than half of the clones encoded an HMG box identical to the published HMG box sequence of *Sox-4*, which was cloned from a mouse embryo cDNA library by low stringency probing with *SRY* (Gubbay *et al.*, 1990). Six other *Sox* sequences were identified with much lower frequencies in the PCR experiment. Five of these sequences



		BINDING SITE	BINDING
1	MWε-1	-GAG AACAAAG GCG- -CTC TTGTTTC CGC-	+++
2	MWε-I12	-GAG IICAAAG GCG- -CTC CCGTTTC CGC-	+++
3	MWε-G12	-GAG GGCAAAG GCG- -CTC CCGTTTC CGC-	-
4	MWε-I456	-GAG AACIIIG GCG- -CTC TTGCCCC CGC-	+++
5	MWε-G456	-GAG AACGGGG GCG- -CTC TTGCCCC CGC-	-
6	MWε-I3	-GAG AACAAAG GCG- -CTC TTITTTC CGC-	+
7	MWε-I7	-GAG AACAAAI GCG- -CTC TTGTTTC CGC-	+
8	MWTα2	-AGC TTCAAAG GGT- -TCG AAGTTTC CCA-	-
9	MWTα2IC	-AGC IICCCCG GGT- -TCG CCGIIIC CCA-	-
10	MWTα2GC	-AGC GGCCCCG GGT- -TCG CCGGGGC CCA-	-
11	MWε-T12	-CAG TTCAAAG GCG- -GTC AAGTTTC CGC-	-

Fig. 2. Sequence-specific binding by the Sox-4 HMG box to the AACAAAG motif. (**A**) Compilation of probes and measured Sox-4 binding in gel retardation analysis. Recombinant Sox-4 HMG box polypeptide was produced in *E.coli* as described in Materials and methods. A strong shift is indicated by +++; a weak shift by +. Full sequences of retardation probes are given in Materials and methods. (**B**) Gel retardation analysis as summarized above. Numbering of lanes corresponds to probe numbers in panel A. Free probes (FP) migrate at different positions due to size differences. Fat arrows indicate strong shifts (lanes 1, 2 and 4); a thin arrow marks weak shifts (lanes 6 and 7). (**C**) Methylation interference footprinting of the Sox-4 HMG box as measured on the positive strand of the AACAAAG motif. The binding motif is indicated with a solid bar on the left. F, free probe; B, bound probe. Fat arrows indicate methylated A residues interfering with binding. No footprint was apparent on the negative strand (not shown), in line with results obtained with recombinant SRY and TCF-1 (van de Wetering *et al.*, 1991; van de Wetering and Clevers, 1992). (**D**) Determination of the equilibrium dissociation constant of Sox-4–MWε-1 complexes by Scatchard analysis. The K_d equals $\sim 3 \times 10^{-11}$ M. Experiment and calculations are described in Materials and methods.

encoded peptides identical to the published sequences of *Sox-1* to *-3* (Gubbay *et al.*, 1990), *Sox-11* and *Sox-14* (Wright *et al.*, 1993). One sequence was novel and was termed *Sox-15* (van de Wetering and Clevers, 1993).

Screening of a mouse thymus lambda-ZAP cDNA library with each of the seven *Sox* sequences yielded only *Sox-4* cDNA clones. As no signals were obtained with any of the other clones, they most likely derived from contaminating genomic DNA in our cDNA preparations. Indeed, the HMG box of *SRY* is not interrupted by introns and would be amplified by our primers (Gubbay *et al.*, 1990; Sinclair *et al.*, 1990). Restriction mapping of 19 independent *Sox-4* cDNA clones did not reveal evidence for alternative splicing. The primary sequence of the longest clone, pSox-4, was then determined (Figure 1A). We detected a single long open reading frame of 440 amino acids starting at bp 659 and extending to bp 1974, with an in-frame stop codon at bp 292. All cDNA clones terminated in an A-rich stretch, which did not represent a polyadenylated tail, as a typical polyadenylation signal was absent and other bases appeared within the A stretch. The encoded protein contained an HMG box located close to the N terminus. Its most striking feature was an extremely serine-rich stretch near the C terminus (Figure 1B). Database searches (EMBL, GenBank) with the full-length protein revealed a strong homology to various other *Sox* genes in the HMG box region; no significant similarities were detected in these data bases with other regions of the Sox-4 protein. Subsequent isolation of a full-length human *Sox-4* cDNA clone revealed high homology towards its mouse orthologue (Figure 1B) and identity to an independently cloned human *Sox-4* cDNA (C.Farr and P.Goodfellow, personal communication).

Sequence-specific DNA binding by Sox-4

We next analysed if Sox-4 could bind to the AACAAAG motif. To that end, the HMG box was excised from pSox-4

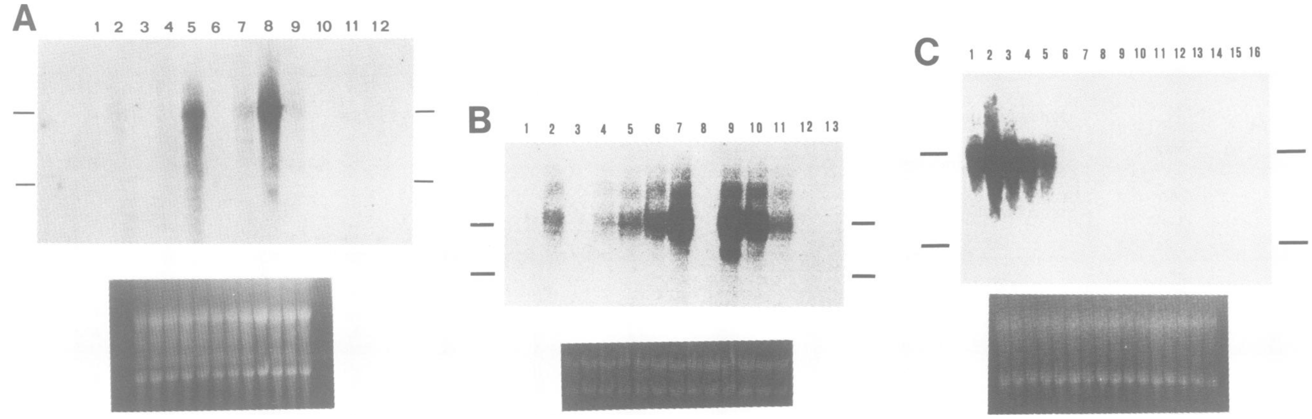


Fig. 3. Tissue-specific expression of *Sox-4* mRNA. (**A**) *Sox-4* mRNA as measured by Northern blotting on whole-tissue RNA. Ribosomal RNAs are indicated with horizontal lines. *Sox-4* mRNA runs as an ~5 kb species. The smears result from deliberate overexposure. Lane 1, spleen; lane 2, lymph node; lane 3, liver; lane 4, kidney; lane 5, ovaries (similar signals were obtained with testes); lane 6, gut; lane 7, lung; lane 8, thymus; lane 9, heart; lane 10, muscle; lane 11, salivary gland; lane 12, brain. Approximately equal amounts of RNA were loaded in each lane as evidenced by inspection of the ethidium bromide-stained gel (lower panel). (**B**) *Sox-4* mRNA as measured by Northern blotting on total RNA extracted from various murine cell lines. Prothymocytes: lane 1, 34.1E and lane 3, 35.1E; Thymocytes/T cells: lane 2, 34.1L; lane 4, 35.1L; lane 5, A1; lane 6, B2; lane 7, EL-4; lane 8, BW5147. Pre-B cells: lane 9, 38B9; lane 10, 40E1; lane 11, 1881. B cells: lane 12, Ag8; lane 13, NS-1. Equal amounts of RNA were present in each lane as judged by examination of the ethidium bromide-stained gel (lower panel). (**C**) Northern blot analysis of *SOX-4* mRNA on a panel of human lymphoid cell lines. Lanes 1–3 are T lineage cells (Jurkat, Molt-4 and Peer, respectively). Lanes 4 and 5 are early pre-B cells (REH and Nalm-6). All other lanes are B lineage cells: lane 6, SMS-SB; lane 7, BJAB; lane 8, Raji; lane 9, MTLM-4; lane 10, APD; lane 11, BSM; lane 12, CRL 1484; lane 13, Fravel; lane 14, U266; lane 15, ARH 77; lane 16, RPMI 8226. Equal amounts of RNA were analysed as is evident from the ethidium bromide-stained gel (lower panel).

by PCR, cloned into the expression vector pET3a (Studier *et al.*, 1990) and expressed in *Escherichia coli*. As analysed by gel retardation, the Sox-4 HMG box indeed bound to the AACAAAG motif (probe MWϵ-1; Figure 2B, lane 1). As described for other HMG boxes, Sox-4 interacted with DNA bases within the minor groove: substitution of A/T pairs for I/C pairs, which leaves the surface of the minor groove intact (Starr and Hawley, 1991), had no apparent effect on binding affinity (lanes 2 and 4). By contrast, substituting A/T for G/C pairs abolished binding (lanes 3 and 5). Interestingly, we did not detect binding of Sox-4 to the TTCAAAG motif of the TCR-α enhancer (lane 8), even when in the context of the CD3-ϵ enhancer (lane 11). In contrast, TCF-1 and human SRY bind with comparable affinities to both motifs (van de Wetering and Clevers, 1992). This indicates that differences in fine specificity exist between related HMG box proteins such as SRY and Sox-4.

Thus far, interactions of HMG boxes with their cognate motifs have only been studied for the A/T pairs at positions 1, 2, 4, 5 and 6 (Giese *et al.*, 1992; van de Wetering and Clevers, 1992). To investigate contacts in the minor groove of the G/C pairs at positions 3 and 7 of the motif, we adapted the original Starr and Hawley A for I substitution (Starr and Hawley, 1991) by replacing G/C pairs for I/C pairs at these positions. This novel substitution leaves the major groove unchanged, but results in the removal of an amine group from the G residue in the minor groove. These substitutions each interfered with binding, establishing that base recognition occurred in the minor groove at all seven positions of the AACAAAG motif (Figure 2B, lanes 6 and 7).

Methylation interference footprinting confirmed the minor groove interactions of the Sox-4 HMG box with the AACAAAG motif of the CD3-ϵ enhancer (Figure 2C). Dimethylsulphate (DMS) methylates A residues in the minor groove on N3, whereas G residues are methylated in the major groove. The footprints were indistinguishable from those obtained with TCF-1 and SRY (van de Wetering *et al.*, 1991; van de Wetering and Clevers, 1992).

The K_d of the interaction of the recombinant Sox-4 HMG box with its AACAAAG motif was determined by Scatchard analysis at $\sim 3 \times 10^{-11}$ M (Figure 2D). The K_d was thus 30-fold lower than that determined for the interaction of LEF-1 and TCF-1 with their cognate motifs ($\sim 10^{-9}$ M; Giese *et al.*, 1991; M.van de Wetering and H.Clevers, unpublished) and was actually within the range determined for 'classical' transcription factors from the other families (Johnson and McKnight, 1989).

Sox-4, like LEF-1, is preferentially expressed in T and pre-B lymphocytes

In order to study whether Sox-4 could be responsible for the observed T cell-specific enhancer effect of the AACAAAG motif, we next determined the tissue-specific expression pattern of the gene by Northern analysis. A major hybridizing band with an apparent size of 5 kb was observed, as well as several minor bands of different sizes on all Northern blots. The nature of these bands has not yet been determined. Significantly, the *Sox-4* gene consists of a single exon (Schilham *et al.*, 1993), suggesting the occurrence of alternative polyadenylation. A deliberate overexposure of the blot revealed high levels of *Sox-4* mRNA in the thymus and gonads of adult mice (Figure 3A, lanes 8 and 5). Much lower levels, only obvious after prolonged exposure, were seen in lymph node, lung and heart (lanes 2, 7 and 9). Densitometric scanning revealed that these levels were at least 50-fold lower than those observed in thymus and gonads. Except for the expression in lymph node, probably caused by the presence of T cells, the significance of this low level expression remains to be determined. In addition, we found moderate expression of Sox-4 in whole brain RNA of neonatal mice. This signal decreased rapidly with age (not shown). To define more precisely the onset of expression in the T lineage, we performed Northern analysis on a small panel of well-defined murine lymphoid cell lines. We thus found Sox-4 to be expressed in mature T cells and in pro/pre-B cells. No expression was observed in

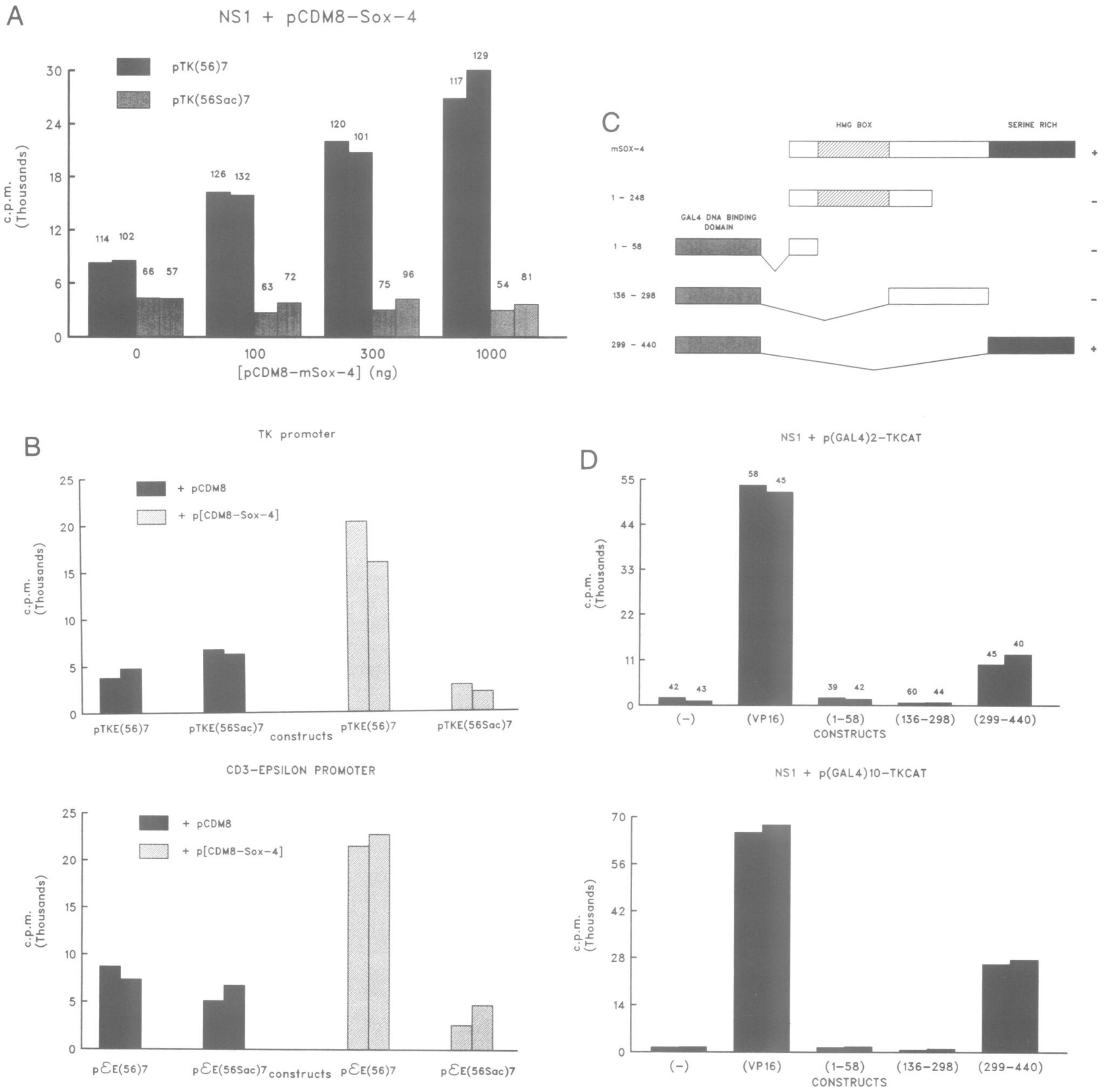


Fig. 4. Transactivation by Sox-4 in a transient CAT assay. (**A**) Co-transfection of *Sox-4* with an AACAAAG concatamer in a reporter–CAT construct leads to an increase in the amount of recovered CAT activity. Transfected cells were the Sox-4 negative B cell line NS-1. CAT reporter constructs: pTK(56)$_7$ refers to the reporter construct which carries seven copies of the MWϵ-1 motif upstream of the TK promoter. pTK(56Sac)$_7$ refers to the reporter construct which carries seven copies of a mutant MWϵ-1 motif. Co-transfection was performed with an increasing amount of the *Sox-4* expression vector pCDM8-Sox-4 as indicated. The amount of expression vector was kept constant at 1 μg by addition of pCDM8. Transfections were performed in duplicate; the duplicate values are given as c.p.m. of extractable [^{14}C]butyryl-chloramphenicol. As a control for transfection efficiency, transfected plasmids were recovered by quantitative Hirt extraction at the time of assay, and transformed into *E.coli* strand DH5α. This strain does not support antibiotic resistance from pCDM8-based vectors; this assay therefore specifically measures the amount of reporter CAT construct. Colony numbers are given over each bar. (**B**) The Sox-4 protein transactivates transcription from a remote position. Co-transfection of *Sox-4* with an AACAAAG concatamer inserted downstream of the CAT reading frame driven by the TK or the CD3-ϵ promoter leads to an increase in the amount of recovered CAT activity. Transfected cells were the B cell line NS-1. CAT reporter constructs: pTKE(56)$_7$ refers to the reporter construct which carries seven copies of the MWϵ-1 motif downstream of the TK/CAT gene; pTKE(56Sac)$_7$ is similar, but carries seven copies of a mutant MWϵ-1 motif. pϵE(56)$_7$ refers to the reporter construct which carries seven copies of the MWϵ-1 motif downstream of the CAT gene driven by the minimal CD3-ϵ promoter. pϵE(56Sac)$_7$ is similar, but carries seven copies of a mutant MWϵ-1 motif. Co-transfections with pCDM8-Sox-4 or with pCDM8 (1 μg per transfection) were performed in duplicate; the duplicate values are given as c.p.m. of extractable [^{14}C]butyryl-chloramphenicol. (**C**) Graphic representation of the expression constructs used in the CAT assays. The numbers in front of the constructs refer to amino acid positions in murine Sox-4. The presence or absence of transactivation by co-transfection with relevant reporter CAT plasmids is summarized behind each construct with + or –. (**D**) Transactivation mediated by chimeric GAL4/Sox-4 proteins on *GAL4/CAT* reporter constructs. NS-1 B cells were electroporated with 20 μg of the reporter–CAT construct containing two GAL4 binding motifs [p(GAL4)$_2$-TKCAT, top] or 10 binding motifs [p(GAL4)$_{10}$-TKCAT, bottom) and 0.2 μg of *GAL4/Sox-4* expression construct. Expression constructs: (–) GAL4-DNA binding domain as negative control; (VP16) GAL4/VP16 chimera as positive control; (1–58), (136–298) and (299–440), see (**B**). Results of Hirt controls [performed only for p(GAL4)$_2$-TKCAT] are given over the bars.

CD3-δ/CD3-ϵ^-, TCR rearrangement$^-$ precursor T cells (prothymocytes), nor in B cells (Figure 3B). Importantly, *Sox-4* mRNA was absent from the T cell line BW5147 in which the AACAAAG enhancer motif is non-functional (Figure 3B, lane 8; see Introduction). Northern blot analysis of a panel of human cell lines revealed a similar distribution, i.e. expression in T and in early pre-B cell lines (Figure 3C). The observed expression pattern in the lymphoid lineage was similar to that reported for LEF-1 (Travis *et al.*, 1991). By contrast, TCF-1 is uniquely expressed in the T lineage (Oosterwegel *et al.*, 1993).

Sox-4 transactivates transcription through an AACAAAG concatamer

We then sought to establish whether Sox-4 could transactivate transcription through an AACAAAG concatamer. To that end, a full-length *Sox-4* cDNA was inserted into the eukaryotic expression vector pCDM8. pCDM8-Sox-4 was co-transfected with pTK(56)$_7$, a reporter−CAT plasmid containing seven copies of the AACAAAG motif inserted upstream of a minimal herpes simplex thymidine kinase (TK) promoter (van de Wetering *et al.*, 1991). As a control, we used pTK(56Sac)$_7$, a TK−CAT vector containing seven copies of an oligonucleotide in which the AACAAAG motif had been replaced by CCGCGGT (van de Wetering *et al.*, 1991). CAT transfections were performed in duplicate and assayed by organic phase separation (van de Wetering *et al.*, 1991).

As shown in Figure 4, Sox-4 transactivated transcription through the AACAAAG concatamer in the murine B lineage myeloma line NS1. Notably, the same transactivation effect was observed upon co-transfection in the Sox-4 negative murine T cell line BW5147 and in the human erythroleukaemic cell line K562 (not shown). To test whether Sox-4 could transactivate transcription from a remote position, we inserted the AACAAAG concatamer (and its mutant CCGCGGT concatamer) downstream of the CAT reading frame. Again, transactivation was observed in NS1 cells (see Figure 4B, upper panel), indicating that Sox-4 can function as an enhancer activator. Similar observations were made with CD3-ϵ promoter−CAT constructs containing the pertinent concatamers inserted downstream of the CAT reading frame (Figure 4B, lower panel), or upstream of the promoter (not shown). Thus, the observed transactivation was not restricted to the TK promoter, but also occurred when the AACAAAG concatamer from the CD3-ϵ enhancer was combined with the minimal CD3-ϵ promoter.

Truncation of the Sox-4 protein at amino acid position 219, effectively removing the serine-rich C terminus, abrogated transactivation (summarized in Figure 4C). We then constructed chimeric proteins consisting of non-HMG box regions from Sox-4 grafted onto the GAL4 DNA-binding domain (Figure 4C). Co-transfection of these chimeras with a GAL4−CAT reporter construct mapped the transactivation domain to the serine-rich C terminus (residues 299−440; Figure 4D). A similar serine-rich transactivation domain has been described previously in the C terminus of the NF-κB p65 subunit (Schmitz and Bauerle, 1991). Importantly, the HMG box region was dispensable for the observed transactivation phenomenon. Co-transfection of *Sox-4* constructs with reporter genes containing a single AACAAAG motif or containing the CD3-ϵ enhancer did not result in transactivation (not shown).

We thus conclude that (i) Sox-4 is expressed in mature T cells and in some pre-B cells, (ii) as described for 'classical' transcription factors (Ptashne, 1988), the Sox-4 protein is modular, as its DNA-binding and transactivation domains are separable, and (iii) Sox-4 most likely mediates the enhancer effect of the AACAAAG motif in lymphocytes.

Discussion

In the present study, we identify the protein encoded by the murine *Sox-4* gene as a transcriptional activator that is capable of binding to the T cell enhancer motif AACAAAG. A *Sox-4* fragment was first obtained by PCR based on homologies existing between HMG boxes of several genes closely related to the mammalian sex determinator *SRY*. Full-length *Sox-4* cDNA clones were subsequently isolated. *Sox-4* mRNA was primarily detected in thymus, gonads and fetal brain. Recombinant *Sox-4* HMG box protein was shown to bind to and transactivate transcription through the AACAAAG motif, contacting all seven base pairs in the minor groove. Our observations solve an apparent contradiction emerging from previous studies. These studies had identified TCF-1 and LEF-1 as proteins with a tissue distribution that coincided with the tissue specificity of the enhancers studied. However, co-transfection of *TCF-1* and/or *LEF-1* with concatamers of their minimal cognate motifs did not result in transcriptional enhancement in non-T cells. Moreover, the T lineage cell line BW5147 did express both TCF-1 and LEF-1, but failed to support enhancer activity of the AACAAAG concatamer. Based on the co-transfection experiments presented here, we propose that the Sox-4 protein (and not TCF-1 and/or LEF-1) is responsible for the observed enhancer activity of the concatamer. This notion is corroborated by the finding that BW5147 cells do not express *Sox-4* mRNA.

Sox-4 is thus the first member of the *Sox* gene family demonstrated to be a transactivator of transcription. These observations might have implications for other members of this large family, including *SRY* itself, as well as for the two fungal mating type genes *matMc* and *mat-A1* (Kelly *et al.*, 1988; Staben and Yanofsky, 1990). In analogy with Sox-4, these regulators of sexual differentiation may exert their pronounced biological effects by controlling transcription of downstream genes.

Two recent studies have stressed the importance of the intriguing DNA-binding properties of sequence-specific HMG boxes. First, the HMG box of LEF-1 reportedly induced a dramatic bend in the DNA helix containing its cognate motif (Giese *et al.*, 1992). Based on this observation, LEF-1 was proposed to function as an architectural element organizing enhancer structure in space. This notion explains why LEF-1 is not a transcriptional activator by itself, but can only increase the activity of the TCR-α enhancer when other transcription factors are bound. Second, a curious observation was made recently on the HMG box of *SRY* (Ferrari *et al.*, 1992). The latter study confirmed the previously reported affinity of *SRY* for the CD3-ϵ enhancer motif. Like HMG-1, however, the *SRY* HMG box appeared to be able to interact specifically with four-way junctions irrespective of sequence. However, it is difficult to reconcile the specific biological functions of *SRY* with an affinity for four-way junctions. It appears more plausible that its function results from the ability to recognize linear DNA motifs in *cis*-acting sequences. Significantly, the present study indicates that the HMG box of Sox-4 mediates DNA binding but is dispensable for subsequent transcription transactivation.

Is Sox-4 involved in the establishment or maintenance of the mature T cell phenotype? Although the Sox-4 protein is predicted to bind with high affinity to the T cell-specific enhancers of the CD3-ε and CD4 genes, no direct evidence is available that would identify these or other T cell genes as *in vivo* targets. Co-transfection of *Sox-4* with the CD3-ε enhancer did not result in transactivation in non-T cells. This was not surprising as the CD3-ε harbours a powerful non-T cell silencer which maps upstream of the AACAAAG motif (M.Oosterwegel and H.Clevers, in preparation). We have found in the present study that at least two Sox-4 transactivation domains need to occupy a regulatory element to mediate transactivation in the absence of other enhancer elements. This observation implies that a single Sox-4 molecule will not drive tissue-specific transcription, but that it probably cooperates with other factors to control gene expression. A cooperative mechanism of tissue-specific gene control was proposed several years ago (Maniatis *et al.*, 1987; Atchison, 1988; Johnson and McKnight, 1989).

Our observations establish that at least three HMG box factors are expressed simultaneously in mature T cells. TCF-1 (Oosterwegel *et al.*, 1991a; van de Wetering *et al.*, 1991) and LEF-1 (Travis *et al.*, 1991) have been extensively described elsewhere, with particular emphasis on their DNA binding properties. Despite the fact that they bind DNA in a specific fashion, these highly homologous factors do not act as classical transcription factors *sensu stricto*. The virtually complete conservation of protein sequences outside the HMG boxes of the human and murine homologues of LEF-1 and TCF-1 (Oosterwegel *et al.*, 1991a; Travis *et al.*, 1991; Waterman *et al.*, 1991) suggests that these proteins perform additional functions. So far, biological functions other than those resulting from DNA binding and bending have not been assigned to either of the two proteins. Unlike LEF-1 and TCF-1, Sox-4 behaves as a classical transcriptional activator The preferential expression of Sox-4 in mature T cells combined with its transactivating potential render it a mechanistically attractive regulator of T cell differentiation. Nevertheless, our observations indicate that a detailed understanding of lymphoid differentiation must integrate the concerted activity of at least three tissue-specific HMG box genes. The challenge now will be to unravel the individual roles of Sox-4, TCF-1 and LEF-1 in lymphoid differentiation.

Materials and methods

Cloning of Sox homologies by guess-mer PCR

Total RNA was isolated from the murine T cell line 34.1 (kindly provided by Dr A.Kruisbeek) using RNAzol according to the manufacturer's instructions (Cinna-Biotecx). One microgram of total RNA was reverse transcribed using oligo(dT) primer and AMV reverse transcriptase at 42°C for 30 min. PCR was performed using oligonucleotides 5′-GGGGAATTCATGGA(TC)GC(GATC)TT(TC)AT(GATC)GT(GATC)TGG-3′ and 5′-GGGAAGCTT(GATC)GG(GATC)CG(AG)TA(CT)TT(GA)TA(GA)T(TC)(GATC)GG-3′ in 50 μl reaction volumes containing 2 U VENT of polymerase according to the manufacturer's procedures (New England Biolabs). Cycle conditions were 90 s at 95°C, 90 s at 45°C, 60 s at 72°C with a 10 min final extension at 72°C after 40 cycles. The reaction products were purified by gel electrophoresis, cut with *Eco*RI and *Hin*dIII and ligated into *Eco*RI/*Hin*dIII-digested pBluescriptSK. After transformation into DH5α, the bacteria were grown on agar plates containing isopropylthiogalactoside (IPTG) and X-Gal. Blue colonies were picked and plasmids were screened for *Sox*-related inserts by sequencing.

Gel retardation assay

Annealed oligonucleotides were labelled by T4 kinase with [γ -^{32}P]ATP. All probes were purified by non-denaturing polyacrylamide electrophoresis. For a typical binding reaction, 1 μl bacterial lysate was 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., equalling 0.2 ng) was added and the mixture was incubated for an additional 20 min. The samples were then electrophoresed through a non-denaturing 8% polyacrylamide gel run in 0.25 × TBE at room temperature.

For the determination of the K_d by Scatchard analysis, as described elsewhere by Verrijzer *et al.* (1990), a fixed amount of *E.coli* extract containing the recombinant HMG box was incubated with the retardation probe MEε-1 at 5.8×10^{-11} M, 7.6×10^{-11} M, 1.5×10^{-10} M, 1.7×10^{-10} M, 3.5×10^{-10} M and 3.6×10^{-10} M under standard conditions. After gel retardation, activity in free (D_f) and bound (D_b) DNA was determined on a PhosphorImager (Molecular Dynamics) allowing the calculation of absolute amounts of DNA from total input DNA. In a bimolecular reaction, the equilibrium dissociation constant (K_d) is given by the equation $1/K_s = D_b/D_f(P_t - D_b)$. For Scatchard analysis this equation is rearranged into $D_b/D_f = (P_t - D_f)/K_d$. The K_d value is obtained from the slope of the plot D_b/D_f versus D_b.

Oligonucleotides used were: MWε-1: GGGAGACTGAGAACAAAGCGCTCTCACAC annealed to CCCGTGTGAGAGCGCTTTGTTCTCAGTCT. MWε-I12: ACTGAGIICAAAGCGCTCT annealed to AGAGCGCTTTGCCCTCAGT. MWε-G12: ACTGAGGGCAAAGCGCTCT annealed to AGAGCGCTTTGCCCTCAGT. MWε-I456: ACTGAGAACIIIGCGCTCT annealed to AGAGCGCCCCGTTCTCAGT. MWε-G456: ACTGAGAACGGGGCGCTCT annealed to AGAGCGCCCCGTTCTCAGT. MWε-I3: GGGAGACTGAGAACAAAGCGCTCTCACAC annealed to AGAGCGCTTTITTCTCAGT. MWε-I7: ACTGAGAACAAAICGCTCT annealed to CCCGTGTGAGAGCGCTTTGTTCTCAGTCT. MWTα2: CCCAGAGCTTCAAAGGGTGCCCTACTTG annealed to GGGCAAGTAGGGCACCCTTTGAAGCTCT. MWTα2IC: CCCAGAGCCCCIIIGGGTGCCCTACTTG annealed to GGGCAAGTAGGGCACCCCCCGIIGCTCT. MWTα2GC: CCCAGAGCCCCGGGGGGTGCCCTACTTG annealed to GGGCAAGTAGGGCACCCCCCGGGGCTCT. MWε-T12: ACTGAGTTCAAAGCGCTCT annealed to AGAGCGCTTTGAACTCAGT. All oligonucleotides were synthesized on an Applied Biosystems Inc. 381A machine.

Methylation interference footprinting

Probes were labelled either at the positive or the negative strand oligonucleotide with [γ-^{32}P]ATP using T4 polynucleotide kinase and purified on a sequencing gel. After annealing, the probes were purified on a non-denaturing acrylamide gel. The labelled probes were partially methylated at purine residues using DMS (Siebenlist and Gilbert, 1980). 100 000 c.p.m. of methylated probe was used in a 5-fold scale-up of the gel retardation binding reaction. After fractionation by gel retardation, 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 reaction products were analysed on a 12.5% polyacrylamide–8 M urea sequencing gel.

Production of recombinant protein in E.coli

The *Sox-4* HMG box was cloned by PCR from pSox-4 DNA using the primers 5′-ATACATATGGCTAAGACGCCCAGTGGCCAC-3′ and 5′-CCCGGATCCTACGACTTCACCTTCTTTCG-3′ and inserted between the *Nde*I and *Bam*HI sites of pET-3. The recombinant protein was produced in the bacterial strain BL21(DE3), after induction by IPTG (0.3 mM) (Studier *et al.*, 1990). After 2 h of induction, cells were collected by centrifugation at 3000 r.p.m. for 20 min and resuspended in 0.1 vol buffer containing 10 mM Na_3PO_4 and 15 mM NaCl. After lysis by sonication, bacterial debris was removed by centrifugation. The supernatant was diluted 100-fold in 10 mM Na_3PO_4, 15 mM NaCl and stored at −70°C.

Cell lines

Murine prothymocytes 34.1E, 35.1E and mature T cell lines 34.1L, 35.1L, A1, B2 (all kindly provided by Dr A.Kruisbeek), EL-4, BW5147; the murine pro/pre-B lineage cell lines 38B9, 1881 and 40E1 and mature B cells Ag-8 and NS-1; human T cells Jurkat, Molt-4, Peer; human pre-B cells REH, Nalm6, SMS.SB and BJAB, and mature B lineage cells Raji, MTLM-4, APD, BSM, CRL 1484, Fravel, U266, ARH 77 and RPMI 8226 were all grown in RPMI 1640 supplemented with 5% fetal calf serum and antibiotics.

Isolation of cDNA clones

Full-length murine *Sox-4* cDNA clones were isolated from the murine thymus cDNA library in lambda-ZAP, kindly provided by Dr A.Berns (10^6

primary recombinant phages; average insert size 1.2 kb) by standard hybridization screening at high stringency using the *Sox-4* PCR fragment as probe. Human *SOX-4* cDNA clones were isolated from an HPB-ALL (T cell) library in pMNC7 (a kind gift from Dr B.Seed) using a PCR fragment spanning bp 658 to 830 of mSox-4 as a probe.

Northern blotting

Total RNA was prepared by cell lysis in 3 M LiCl−6 M urea. After precipitation, RNA was phenol−chloroform and chloroform extracted twice and ethanol precipitated. 10 μg per lane of total RNA were run for Northern analysis on 1% agarose containing 3% formaldehyde. RNA was transferred to nitrocellulose and hybridized with *Sox-4* cDNA probes, labelled by random oligo priming, all according to standard procedures (Sambrook *et al.*, 1989).

CAT assays

Cells were transfected by electroporation. In short, 10×10^6 cells were transiently transfected with 20 μg of CAT reporter plasmid and 1 μg of pCDM8/pCDM8-SOX-4 or 200 ng of GAL4 expression plasmid in a volume of 250 μl. Pulse conditions were 960 μF and 250 V using a Gene Pulser Apparatus (Bio-Rad). Cells were harvested 48 h later and freeze−thawed in 100 μl of 100 mM NaCl, 10 mM Tris, pH 7.4, 1 mM EDTA, 50 μl of the lysate were 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 plasmids at the time of the assay, a quantitative Hirt extraction was performed on one-tenth of the cells. Cells were pelleted, resuspended in 0.4 ml of 0.6% SDS−1 mM EDTA and incubated for 20 min at room temperature. Subsequently, 0.01 ml of 5 M NaCl was added and after 12 h at 4°C the samples were spun (5 min at 15 000 r.p.m.). The supernatants were phenol-extracted after which 3 μl were transformed into DH-5α. Plasmids: pCDM8-Sox-4 was constructed by subcloning a *Hin*dIII−*Eco*RV (165−2075) fragment into the *Hin*dIII and *Xho*I (blunted) sites of pCDM8. pCDM8-Sox-4 (1−248) was constructed by digesting pCDM8-Sox-4 with *Bst*XI followed by blunting and religation. *GAL4−Sox-4* chimeras were constructed as follows. pJ3 (1−58), derived by subcloning a PCR fragment spanning amino acids 1−58 into *Bam*HI/*Eco*RI-digested pJ3Ω [an expression vector containing the GAL4 DNA-binding domain (amino acids 1−147), a kind gift of Dr R.Bernards] (primers used: TTTGGATCCATGGTACAACAGACC and TTTGAATTCAGTGGCCACTGGGCGTCTT); pJ3 (136−298), derived by subcloning a PCR fragment spanning amino acids 136−298 into *Bam*HI/*Eco*RI-digested pJ3Ω (primers used: TTTGGATCCGGCAACGCGGGCGCG and TTTGAATTCACGGGCCTCCATCTTCGT); pJ3 (299−440), derived by subcloning a *Sma*I−*Kpn*I fragment from pSox-4 into *Sma*I/*Kpn*I-digested pJ3Ω. pTK(56)$_7$ and pTK(Sac)$_7$ CAT reporter constructs are described in detail by van de Wetering *et al.* (1991). p(GAL4)$_2$-TKCAT and p(GAL4)$_{10}$-TKCAT were constructed by ligating annealed oligonucleotides GATCGGAAGACTCTCCTCC and GATCGGAGGAGAGTCTTCC into *Bam*HI-digested pBLCAT$_2$. The number of GAL4 binding sites was determined by sequencing. pTEK(56)$_7$, derived from pTKCAT by inserting a *Hin*dIII−*Bam*HI fragment containing seven copies of the MWϵ-1 motif from pTK(56)$_7$ into the *Sma*I site downstream of the CAT gene; pTKE(56Sac)$_7$, derived from pTKCAT by inserting a *Hin*dIII−*Bam*HI fragment containing seven copies of the mutated MWϵ-1 motif from pTK(56Sac)$_7$ into the *Sma*I site downstream of the CAT gene; pϵE(56)$_7$, derived by inserting a *Hin*dIII−*Bam*HI fragment containing seven copies of the MWϵ-1 motif from pTK(56)$_7$ into the *Sma*I site downstream of the CAT gene driven by the minimal CD3-ϵ promoter (Clevers *et al.*, 1989); pϵE(56Sac)$_7$, derived by inserting a *Hin*dIII−*Bam*HI fragment containing seven copies of the mutated MWϵ-1 motif from pTK(56Sac)$_7$ into the *Sma*I site downstream of the CAT gene driven by the minimal CD3-ϵ promoter (Clevers *et al.*, 1989).

Acknowledgements

The authors wish to thank Drs J.van de Winkel, T.Logtenberg and M.Schilham for critically reading the manuscript and the members of the laboratory for helpful discussions and support. H.C. is a holder of a 'Pionier' grant from the Dutch scientific organization NWO.

References

Atchison,M.L. (1988) *Annu. Rev. Cell Biol.*, **4**, 127−153.

Castrop,J., Hoevenagel,R., Young,J.R. and Clevers,H. (1992a) *Eur. J. Immunol.*, **22**, 1327−1330.

Castrop,J., van Norren,K. and Clevers,H. (1992b) *Nucleic Acids Res.*, **20**, 611.

Clevers,H.C., Lonberg,N., Dunlap,S., Lacy,E. and Terhorst,C. (1989) *EMBO J.*, **8**, 2527−2535.

Denny,P., Swift,S., Brand,N., Dabhade,N., Barton,P. and Ashworth,A. (1992a) *Nucleic Acids Res.*, **20**, 2887.

Denny,P., Swift,S., Connor,F. and Ashworth,A. (1992b) *EMBO J.*, **11**, 3705−3715.

Dooijes,D., van de Wetering,M., Knippels,L. and Clevers,H. (1993) *J. Biol. Chem.*, in press.

Ferrari,S., Harley,V.R., Pontiggia,A., Goodfellow,P.N., Lovell-Badge,R. and Bianchi,M.E. (1992) *EMBO J.*, **11**, 4497−4506.

Giese,K., Amsterdam,A. and Grosschedl,R. (1991) *Genes Dev.*, **5**, 2567−2578.

Giese,K., Cox,J. and Grosschedl,R. (1992) *Cell*, **69**, 185−195.

Gubbay,J., Collignon,J., Koopman,P., Capel,B., Economou,A., Münsterberg,A., Vivian,N., Goodfellow,P. and Lovell-Badge,R. (1990) *Nature*, **346**, 245−250.

Harley,V.R., Jackson,D.I., Hextall,P.J., Hawkins,J.R., Berkovitz,G.D., Sockanathan,S., Lovell-Badge,R. and Goodfellow,P.N. (1992) *Science*, **255**, 53−56.

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.

Laudet,V., Stehelin,D. and Clevers,H. (1993) *Nucleic Acids Res.*, **21**, 2493−2501.

Maniatis,T., Goodbourn,S. and Fischer,J.A. (1987) *Science*, **236**, 1237−1245.

Nasrin,N., Buggs,C., Kong,X.F., Carnazza,J., Goebl,M. and Alexander-Bridges,M. (1991) *Nature*, **345**, 317−320.

Oosterwegel,M., van de Wetering,M., Dooijes,D., Klomp,L., Winoto,A., Georgopoulos,K., Meijlink,F. and Clevers,H. (1991a) *J. Exp. Med.*, **173**, 1133−1142.

Oosterwegel,M., van de Wetering,M., Holstege,F., Prosser,H.M., Owen,M.J. and Clevers,H. (1991b) *Int. Immunol.*, **3**, 1189−1192.

Oosterwegel,M., van de Wetering,M., Timmerman,J., Kruisbeek,A., Destree,O., Meijlink,F. and Clevers,H. (1993) *Development*, **118**, 439−448.

Parisi,M.A. and Clayton,D.A. (1991) *Science*, **252**, 965−969.

Ptashne,M. (1988) *Nature*, **335**, 683−689.

Sawada,S. and Littman,D.R. (1991) *Mol. Cell. Biol.*, **11**, 5506−5515.

Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual.* 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Schilham,M.W., van Eijk,M., van de Wetering,M. and Clevers,H. (1993) *Nucleic Acids Res.*, **21**, 2009.

Schmitz,M.L. and Baeuerle,P.A. (1991) *EMBO J.*, **10**, 3805−3817.

Seipel,K., Georgiev,O. and Schaffner,W. (1992) *EMBO J.*, **11**, 4961−4968.

Siebenlist,U. and Gilbert,W. (1980) *Proc. Natl Acad. Sci. USA*, **77**, 122−126.

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.

Staben,C. and Yanofsky,C. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 4917−4921.

Starr,D.B. and Hawley,D.K. (1991) *Cell*, **67**, 1231−1240.

Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1990) *Methods Enzymol.*, **185**, 60−89.

Sugimoto,A., Lino,Y., Maeda,T., Watanabe,Y. and Yamamoto,M. (1991) *Genes Dev.*, **5**, 1990−1999.

Travis,A., Amsterdam,A., Belanger,C. and Grosschedl,R. (1991) *Genes Dev.*, **5**, 880−894.

van de Wetering,M. and Clevers,H. (1992) *EMBO J.*, **11**, 3039−3044.

van de Wetering,M. and Clevers,H. (1993) *Nucleic Acids Res.*, **21**, 1669.

van de Wetering,M., Oosterwegel,M., Dooijes,D and Clevers,H. (1991) *EMBO J.*, **10**, 123−131.

Verrijzer,C.P., Kal,A.J. and van der Vliet,P.J. (1990) *Genes Dev.*, **4**, 1964−1974.

Waterman,M.L., Fischer,W.H. and Jones,K.A. (1991) *Genes Dev.*, **5**, 656−669.

Wen,L., Huang,J.K., Johnson,B.H. and Reeck,J.R. (1989) *Nucleic Acids Res.*, **17**, 1197−1214.

Wright,E.M., Snopek,B. and Koopman,P. (1993) *Nucleic Acids Res.*, **21**, 744.

Received on April 28, 1993; revised on June 30, 1993