The EMBO Journal vol.11 no.10 pp.3705-3712, 1992

An *SRY*-related gene expressed during spermatogenesis in the mouse encodes a sequence-specific DNA-binding protein

Paul Denny, Sally Swift, Frances Connor and Alan Ashworth[1]

Chester Beatty Laboratories, The Institute of Cancer Research, Fulham Rd, London SW3 6JB, UK

[1]Corresponding author

Communicated by R.A.Weiss

***SRY*, the testis determining gene, encodes a member of a family of DNA binding proteins characterized by an amino acid sequence motif known as the HMG box. Using degenerate primers and the polymerase chain reaction, we have isolated *SRY*-related cDNAs from adult murine testis RNA. One of these, *Sox-5*, encodes a 43 kDa HMG-box protein with similarities to transcription activating proteins. Anti-Sox-5 antibody was used to analyse expression of Sox-5 in pre-pubertal testis and in fractionated spermatogenic cells. Sox-5 is restricted to post-meiotic germ cells, being found at highest levels in round spermatids. Sox-5 is a DNA binding protein and binding site selection assays suggest that it can bind specifically to oligonucleotides containing the consensus motif AACAAT. Sry can also bind to this motif, indicating that the Sry family may have overlapping sequence specificities.**

Key words: HMG-box/sex determination/transcription factor

Introduction

Sex in mammals is determined by a dominantly acting gene present on the Y chromosome which diverts the differentiation of the genital ridge from the 'default' ovarian pathway to that of the testis (McLaren, 1988). Recently a gene, *SRY* has been identified in the sex determining regions of the human and mouse Y chromosomes (Gubbay *et al.*, 1990; Sinclair *et al.*, 1990) and proposed as a candidate for this testis determining gene. When stably introduced into the mouse genome, *Sry* directs the formation of testes in XX females, showing that this gene is indeed the testis determining gene (Koopman *et al.*, 1991). *Sry* encodes a protein with a DNA binding motif known as a HMG-box (Jantzen *et al.*, 1990), which is consistent with the cell-autonomous role proposed for the testis determining gene (Burgoyne *et al.*, 1988). This motif has been found in several sequence-specific DNA binding proteins, such as the T cell-specific factors TCF-1 (Van de Wetering *et al.*, 1991) and TCF-1α (Waterman *et al.*, 1991), and the RNA polymerase I transcription factor UBF (Jantzen *et al.*, 1990). The HMG-box motif(s) in these factors are both necessary and sufficient for sequence-specific DNA binding.

Four *Sry*-related genes (*Sox* genes) have been described in the mouse (Gubbay *et al.*, 1990). These are autosomal and share considerable homology over the HMG-box region. *Sry* and these *Sry*-related genes are expressed during embryogenesis (Gubbay *et al.*, 1990) and *Sry* is re-expressed in the adult testis (Koopman *et al.*, 1990). It is probable that Sry functions in the embryo as a transcription factor, regulating effector genes that directly or indirectly determine the path of differentiation taken by the genital ridge. Its function in the adult testis, however, is uncertain and it joins a growing number of embryonically expressed genes that are activated specifically in testis (Rubin *et al.*, 1986; Shackleford and Varmus, 1987; Wolgemuth *et al.*, 1987). We have isolated further *Sry*-like genes from adult mouse testis cDNA (Denny *et al.*, 1992) and here we describe one of these genes, *Sox-5*. We show that *Sox-5* mRNA is at elevated levels in the testis and that Sox-5 protein is present in the post-meiotic germ cells. We also demonstrate that Sox-5 is a DNA binding protein with a sequence specificity overlapping that of Sry.

Results

SRY-related transcripts are present in adult mouse testis

To identify novel *Sry*-related genes expressed in the adult testis, we made use of the polymerase chain reaction (PCR) to amplify *Sry*-related cDNAs. Poly(A)$^+$ RNA reverse-transcribed into cDNA, or phage lysates from cDNA libraries were used as templates for PCR, using mixed-sequence oligonucleotide primers that corresponded to regions conserved in the known *SRY* and *SRY*-related amino acid sequences. Primer sequences were chosen so as to anneal specifically with *SRY*-like sequences and so avoid amplification of other HMG-box genes. For example, the amino acids valine and tryptophan at positions 22 and 23 in Figure 1b, which correspond to the first five nucleotides of the sense strand primer are found only in the *SRY* family and not at equivalent positions in other HMG box sequences.

Specific PCR products of the predicted size were cloned and sequenced. Two distinct mouse cDNAs were isolated from testis, which we have named *Sox-5* and *Sox-6*. These cDNAs are different to those *Sry*-related clones previously described (Gubbay *et al.*, 1990; Denny *et al.*, 1992). *Sry* was not recovered from amplifications performed on testis mRNA, despite the fact that the *Sry* gene is expressed in adult mouse testis (Koopman *et al.*, 1990). This may be a result of the differing levels of expression of *Sry* and the other, *Sry*-related genes or reflect some property of the *Sry* transcript which makes it difficult to reverse transcribe or to amplify its cDNA by PCR. One of these *Sry*-like clones, *Sox-5*, was abundantly expressed in testis and we proceeded to obtain full-length cDNAs.

Sox-5 is a novel *SRY*-related murine gene

The published sequences for members of the *SRY*-related gene family are restricted to the HMG-box region. It was therefore of great interest to obtain sequence information outside this region, as it could give clues to the functions

a

```
                                                                                                   90
GCAGTTCTTATGAAGCCTCTTTCTTCAGCGTACATGATACGTCCTCCCCTGCTCGCCCCAGAGGTGACCCTTACCCTGTTCAGCTGATCC
                                                                                                  180
CAACCACTATGGCAGCTGCTGCTGCAGCAACACCAGGCTTAGGCCCACTCCAGCTGCAGGATGAAGTGGCCCAGCCACTGAACCTATCAG
         M  A  A  A  A  A  A  T  P  G  L  G  P  L  Q  L  Q  D  E  V  A  Q  P  L  N  L  S
                                                                                                  270
CTAAGCCCAAGACCTCTGATGGCAAATCACCTGCATCACCCACCTCTCCTCACATGCCAGCTCTGAGAATAAACAGTGGGGCTGGCCCCC
A  K  P  K  T  S  D  G  K  S  P  A  S  P  T  S  P  H  M  P  A  L  R  I  N  S  G  A  G  P
                                                                                                  360
TCAAAGCCTCTGTCCCGGCAGCATTAGCTAGTCCTTCTGCAAGAGTTAGCACAATAGGCTATTTAAATGACCACGATGCTGTCACCAAGG
L  K  A  S  V  P  A  A  L  A  S  P  S  A  R  V  S  T  I  G  Y  L  N  D  H  D  A  V  T  K
                                                                                                  450
CAATCCAAGAAGCTCGCCAGATGAAAGAGCAACTCAGGCGAGAGCAGCAAGCGTTGGACGGAAGGTGGCCGTGGTGAACAGCATAGGTC
A  I  Q  E  A  R  Q  M  K  E  Q  L  R  R  E  Q  Q  A  L  D  G  K  V  A  V  V  N  S  I  G
                                                                                                  540
TCAGCAACTGCCGGACAGAAAAGGAAAAAACAACACTGGAGAGCCTGACTCAGCAGTTGGCAGTTAAACAGAATGAAGAAGGGAAATTTA
L  S  N  C  R  T  E  K  E  K  T  T  L  E  S  L  T  Q  Q  L  A  V  K  Q  N  E  E  G  K  F
                                                                                                  630
GCCATGGAATGATGGATTTCAATATGAGTGGAGATTCTGACGGAAGCGCTGGAGTCTCAGAGTCAAGAATTTACAGGGAATCCAGGGGAC
S  H  G  M  M  D  F  N  M  S  G  D  S  D  G  S  A  G  V  S  E  S  R  I  Y  R  E  S  R  G
                                                                                                  720
GTGGTAGCAACGAGCCCCACATAAAGCGTCCAATGAATGCCTTCATGGTGTGGGCGAAAGATGAACGGAGGAAAATCCTTCAGGCCTTTC
R  G  S  N  E  P  H  I  K  R  P  M  N  A  F  M  V  W  A  K  D  E  R  R  K  I  L  Q  A  F
                                                                                                  810
CCGACATGCACAATTCCAACATCAGCAAGATACTGGGATCTCGCTGGAAAGCTATGACCAACCTAGAGAAACAGCCATACTATGAGGAGC
P  D  M  H  N  S  N  I  S  K  I  L  G  S  R  W  K  A  M  T  N  L  E  K  Q  P  Y  Y  E  E
                                                                                                  900
AGGCCCGCCTCAGCAAACAGCACCTGGAGAAGTACCCGGACTATAAGTACAAGCCTAGGCCGAAGCGCACCTGTCTGGTGGACGGCAAGA
Q  A  R  L  S  K  Q  H  L  E  K  Y  P  D  Y  K  Y  K  P  R  P  K  R  T  C  L  V  D  G  K
                                                                                                  990
AACTGCGTATCGGGGAGTACAAGGCCATCATGCGGAACCGGAGGCAGGAAATGCGACAGTACTTCAATGTTGGGCAACAAGCACAGATCC
K  L  R  I  G  E  Y  K  A  I  M  R  N  R  R  Q  E  M  R  Q  Y  F  N  V  G  Q  Q  A  Q  I
                                                                                                 1080
CCATCGCGACGGCCGGAGTTGTATACCCCGGCGCCATCGCCATGGCAGGAATGCCGTCCCCTCACCTGCCCTCCGAGCACTCGAGCGTGT
P  I  A  T  A  G  V  V  Y  P  G  A  I  A  M  A  G  M  P  S  P  H  L  P  S  E  H  S  S  V
                                                                                                 1170
CCAGCAGCCCGGAGCCCGGGATGCCCGTGATCCAGAGCACTTACGGCGCCAAGGGAGAGGAGCCCCACATCAAGGAAGAGATCCAGGCTG
S  S  S  P  E  P  G  M  P  V  I  Q  S  T  Y  G  A  K  G  E  E  P  H  I  K  E  E  I  Q  A
                                                                                                 1260
AGGACATCAACGGAGAGATTTACGAGGAGTACGATGAGGAGGAGGAGGACCCGGATGTGGATTATGGGAGTGACAGCGAAAACCACATTG
E  D  I  N  G  E  I  Y  E  E  Y  D  E  E  E  E  D  P  D  V  D  Y  G  S  D  S  E  N  H  I
                                                                  ____________________________
                                                                                                 1350
CGGGACAAGCCAACTGATAAGGGCCAACAGACTGTGGTGAGCCGAGGACTTGAAGAAGCCCTGTCCGGTTCATCCTTCCCAGTGGCCAAG
A  G  Q  A  N
_____________

                                                                                                 1440
CACATTAACTCTCTCATACACTGACTGTTATTTTAACTGTTAGTCTTATATAGTTGGGACATCAGCTGACAAATAGACCTCAGCCTCAAA

                                                                                                 1530
AGGCTCGGAAAGAAAAGAAAGAAAGAAAGAAAAAAAAAAAAAAAAAATTACAAGCAAAAACGTCATCAAGAAGAAGAGATCGAAATAAGCT

                                                                                                 1620
ATGGGTTAAACAGTGCCAGTAACTCAGCTGCTACACCCCAGCACTGAAGTCTCACCCGTCAACTTTTTTTTTTTTTTTAATAAACTTTAT

                        1644
GGCTGTTTGTTCTAAAAAAAAAAA
```

b

```
ATGAATGCTTTTATTGTTTGG 3'                                                 5' GGTTTATATTTATAATTGGG
   C  C  C  G  C                                                            CCGG  C  G  G CA
      A     A  G                                                            G               T
               A                                                            A

Sox-5   PHIKRPMNAF MVWAKDERRK ILQAFPDMHN SNISKILGSR WKAMTNLEKQ PYYEEQARLS KQHLEKYPDY KYKPRPKRT     % homology
              ---- ---                                                            --- ----           100
        11                                                                               89
mSry    GHVKRPMNAF MVWSRGERHK LAQQNPSMQN TEISKQLGCR WKSLTEAEKR PFFQEAQRLK ILHREKYPNY KYQPHRRAK       53
rSry    ERVKRPMNAF MVWSQHQRRQ VALQNPKMRN SDISKQLGHQ WKMLSEAEKW PFFQEAQRLQ AMHKEKYPDY KYRPRRKVK       52
hSRY    DRVKRPMNAF IVWSRDQRRK MALENPRMRN SEISKQLGYQ WKMLTEAEKW PFFQEAQKLQ AMHREKYPNY KYRPRRKAK       52
Sox-1   DRVKRPMNAF MVWSRGQRRK MAQENPKMHN SEISKRLGAE WKVMSEAEKR PFIDEAKRLR ALHMKEHPDY KYRPRRKTK       53
Sox-2   DRVKRPMNAF MVWSRGQRRK MAQENPKMHN SEISKRLGAE WKLLSETEKR PFIDEAKRLR ALHMKEHPDY KYRPRRKTK       53
Sox-3   DRVKRPMNAF MVWSRGQRRK MALENPKMHN SEISKRLGAD WKLLTDAEKR PFTDEAKRLR AVHMKEYPDY KYRPRRKTK       53
Sox-4   GHIKRPMNAF MVWSQIERRK IMEQSPDMHN AEISKRLGKR WKLLKDSDKI PFIQEAERLR LKHMADYPDY KYRPRKKVK       57

Mc      ERTPRPPNAF ILYRKEKHAT LLKSNPSINN SQVSKLVGEM WRNESKEVRM RYFKMSEFYK AQHQKMYPGY KYQPRKNKV       30

TCF-1   PTIKKPLNAF MLYMKEMRAK VIAECTLKES AAINQILGRR WHALSREEQA KYYELARKER QLHMQLYPGW SARDNYGKK       32

HMG-1   NAPKRPPSAF FLFCSEYRPK IKGEHPGLSI GDVAKKLGEM WNNTAADDKQ PYEKKAAKLK EKYEKDIAAY RAKGKPDAA       27
```

Fig. 1. a. *Sox-5* cDNA and deduced amino acid sequence. The sequence was assembled from four overlapping cDNAs, pD578, pD588, pD597 and pD598: only pd598 exhibits any differences from the sequences shown (four differences, at nucleotide positions 854, C for T; 872, A for G; 968, C for T; and at 1020, A for G). One nucleotide difference gives rise to an amino acid change, at nucleotide 1020, substituting serine for glycine. The C-terminal amino acid residues included in the peptide used for immunizations are underlined. **b**. Amino acid sequence of the Sox-5 HMG-box compared with SRY and other HMG-box proteins. The sequences shown are from *Schizosaccharomyces pombe* Mc (Kelly *et al.*, 1988), human TCF-1 (Van de Wetering *et al.*, 1991) and human HMG-1 (Wen *et al.*, 1989). Degenerate oligonucleotides corresponding to highly conserved amino acids (underlined) in the SRY family (Gubbary *et al.*, 1990) were used in the PCR to amplify *SRY*-related cDNAs from mouse RNA. The sequence data reported here have been deposited in the EMBL database under the accession number X65657.

of the proteins encoded by these genes. An adult testis cDNA library was screened with a mixed probe consisting of *Sox-5* and *-6*, which gave numerous positives. Following initial characterization by restriction mapping and partial sequencing, four cDNA clones were shown to overlap and to encode the *Sox-5 SRY*-box sequence (Figure 1a). These clones were sequenced in their entirety. Four nucleotide differences were seen in the overlapping region, one of which gave rise to a difference at the amino acid level, a change from glycine to serine. This level of polymorphism is unusual even though the cDNA library was constructed using mRNA from an outbred strain of mouse. Hybridization of Southern blots with a probe lacking the region encoding the Sry-box gave single bands, suggesting that there is a single *Sox-5* gene, but probes including the box detect an additional band(s) at high stringency (40 mM $NaPO_4$, 1% SDS at 67°C), perhaps due to a gene(s) closely related to *Sox-5* (not shown).

The open reading frame which includes the SRY-box encodes a 392 aa (amino acid) protein with a predicted molecular weight of 43.3 kDa (Figure 1a). A region to the N-terminal side (aa 9−69) of the SRY-box is rich in proline residues (18%) and an overlapping domain (aa 27−74) is rich in serine (20%). There is a single HMG-box located between aa 183−261 and a comparison with other HMG-box domains is shown in Figure 1b. It is clear that Sox-5 is a member of the Sry-related family of HMG-box proteins, as there is a high level of sequence homology with this family, ranging from 52−57%. The C-terminal region of the Sox-5 protein (aa 347−392) is rich in acidic residues (41% glutamic and aspartic acid) and has a high α-helix forming potential (Garnier *et al.*, 1978), making it similar in character to the activation domains of many transcription factors (Jones, 1990). This acidic region is distinct from the polyacidic C-terminal 'tails' found in HMG1 and HMG2, which are simple copolymers of aspartic and glutamic acid (Goodwin and Bustin, 1988). Outside the SRY-box, *Sox-5* is not closely related to sequences in the EMBL or Swiss-Prot databases (release numbers 28 and 19, respectively).

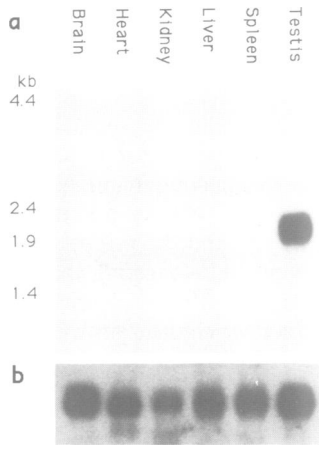

Fig. 2. Northern blot analysis of *Sox-5* expression. **a**. Three micrograms of poly(A)$^+$ RNA extracted from various adult male mouse tissues was electrophoresed, transferred to Hybond N$^+$ (Amersham) and hybridized with a *Sox-5* cDNA probe as described previously (Denny and Ashworth, 1991). **b**. The same filter was subsequently hybridized with an actin cDNA probe as a control for mRNA integrity and loading.

Sox-5 mRNA expression is elevated in testis

We analysed steady-state mRNA levels in adult murine tissues by Northern blot hybridization and found that a 2 kb *Sox-5* mRNA was detected only in poly(A)$^+$ RNA from testis (Figure 2). *Sox-5* mRNA was also not detected in poly(A)$^+$ RNA from skeletal muscle, day 13.5 embryos and a myeloma cell line, P3/NS1/1-Ag4.1 (data not shown). Integrity and equality of loading of the RNAs were confirmed by re-hybridization of the filter with an actin cDNA probe.

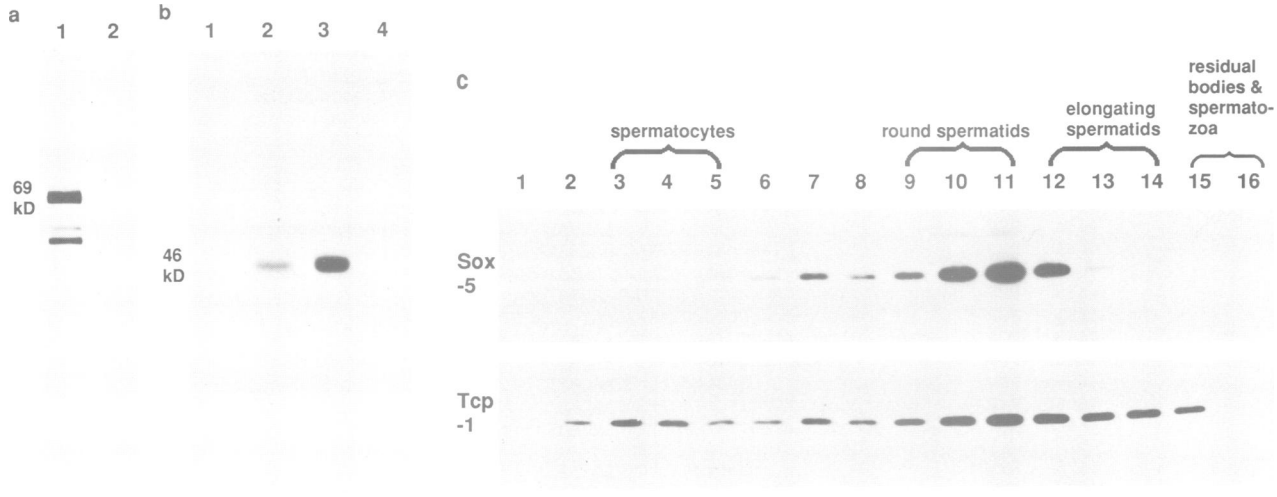

Fig. 3. Immunoblot analysis of *Sox-5* expression. **a**. Total protein extracts from bacteria transformed with a plasmid encoded GST−Sox-5 (lane 1) or GST fused with the *Sox-5* cDNA in the reverse orientation (lane 2). Anti-Sox-5 antibody (137-5) was used at a dilution of 1:20 000. **b**. Total protein extracts from adult mouse spleen (lane 1), testis (lane 2), COS cells transfected with Sox-5 expression plasmid (lane 3) or plasmid lacking insert (lane 4). Sox-5 antibody 137-5 used at dilution of 1:10 000. **c**. Total protein extracts from germ cells fractionated on a CelSep (Dupont) apparatus. Purities of fractions were as follows: fraction 3 contains 85% spermatocytes; fraction 10 contains 90% round spermatids; fraction 13 contains 50% elongating spermatids (main contaminants in this fraction were round spermatids and residual bodies) and fraction 16 contains 80% residual bodies, contaminated mainly with mature spermatozoa. Sox-5 antibody 137-5 was used at a dilution of 1:2000. TCP-1 was detected on the same blot, as a control for the germ-cell fractionation.

Sox-5 is abundant in post-meiotic cells

To allow detection of the *Sox-5* gene product, rabbit polyclonal antisera were raised against a peptide representing the predicted C-terminal 14 aa of Sox-5. Antibodies were affinity purified and shown to react with recombinant glutathione-S-transferase (GST)−Sox-5 fusion protein from *Escherichia coli* (Figure 3a). Sox-5 immunoreactivity is present in extracts from bacteria containing a plasmid expressing the *Sox-5* cDNA in the appropriate orientation, but not when the cDNA is in the reverse orientation. A 46 kDa, testis-specific band is detected on immunoblots of whole-tissue protein extracts (Figure 3b), but not in extracts from a range of adult tissues including brain, heart, kidney, liver and lung (not shown). This is similar to the size predicted for Sox-5 protein from the cDNA sequence and consistent with the observed pattern of expression of *Sox-5* gene mRNA. Furthermore, COS cells transfected with a eukaryotic expression plasmid containing the longest *Sox-5* cDNA synthesize a protein indistinguishable in size from that in whole testis protein extract, demonstrating that this cDNA includes the complete open reading frame of the *Sox-5* gene.

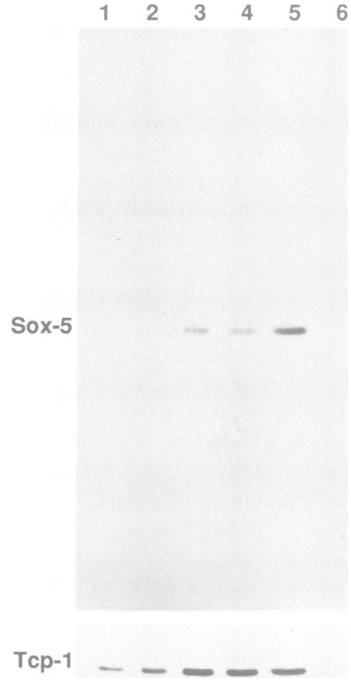

Fig. 4. Immunoblot analysis of expression of Sox-5 in testis from pre-pubertal and adult mice. Total protein extracts from the testes of mice of differing ages analysed as described above. Lane 1, 7 day-old; lane 2, 14 day-old; lane 3, 21 day-old; lane 4, 28 day-old; lane 5, adult (4 months); lane 6, adult spleen. Sox-5 antibody used at a dilution of 1:10 000. TCP-1 was detected on the same blot as a control; an increase in signal is seen between days 14 and 21 as expected (Willison *et al.*, 1989) and signal was detectable from spleen after longer exposure (not shown).

The adult testis contains several distinct cell populations, which can be divided broadly into cells of somatic and germ lineage. Viable germ cells representing the different stages of spermatogenesis can be purified by sedimentation velocity through a bovine serum albumin gradient under unit gravity (Romrell *et al.*, 1976). The main germ cell types are the diploid spermatocytes and the post-meiotic, haploid cells, which are the round and elongating spermatids and mature spermatozoa (Bellve, 1979). Immunoblots of protein extracted from fractionated male germ cells demonstrate that Sox-5 reactivity is highest in the early post-meiotic cells, round spermatids (Figure 3c). An antibody specific for TCP-1, a protein that is present at high levels in the testis (Willison *et al.*, 1989) was used to re-probe the filter and shows the expected peaks of protein in both spermatocytes and spermatids (Willison *et al.*, 1990).

A complementary approach to analysing the expression of *Sox-5* makes use of the first round of spermatogenesis during mouse post-natal testis development. Type A spermatogonia, the cycling stem cells, first appear on day 7 after birth and give rise to committed spermatogenic cells. Meiotic prophase begins during days 9−10 and following

```
4674   FOR                    gaattcgcctcTAACAATGTACTATTGTTGATACATAcgacaggatcc
4674   REV              ggatcctgtcgTATGTATCAACAATAGTACATTGTTAgaggcgaattc
4675   FOR                 gaattcgcctcCACTAAACAATTCAAGCCCGGCTTTAcgacaggatcc
4678   FOR                  gaattcgcctcTAAAAACAATTCACCGATGATTGTTAcgacaggatcc
4678   REV                     ggatcctgtcgTAACAATCATCGGTGAATTGTTTTTAgaggcgaattc
4681   REV                     ...TGTCGGATAAACAATTTAATTTACAT...
4682   FOR                     gaattcgcctcAAACAATACTTATTGTTGACCGCCGCcgacaggatcc
4682   REV            ggatcctgtcgGCGGCGGTCAACAATAAGTATTGTTTgaggcgaattc
4684   REV                                ...AACAATGGGGAACAAACAATAGcgacaggatcc
4684   REV                  ...AACAATGGGGAACAAACAATAGcgacaggatcc
4686   FOR                     gaattcgcctcTAACAATACTGAAGCAATAACTCCACcgacaggatcc
4688   REV  ggatcctgtcgTCATGACAATTTGTCTTGCTAACAATgaggcgaattc
4691   FOR          gaattcgcctcCCCCCCTAAAAGAACAATACTTCCCAcgacaggatcc
4693   FOR     gaattcgcctcGATACTCTACCCCCTATAACAATAGTcgacaggatcc
4696a  FOR               ggatcctgtcgTAGGTATCAACAATAGCACATTGTTAgaggcgaattc
4696a  REV                     gaattcgcctcTAACAATGTGCTATTGTTGATACCTAcgacaggatcc
4696b  FOR      gaattcgcctcATTGTTGCTACTTGATAACAATCTCAcgacaggatcc
4696b  REV  ggatcctgtcgTGAGATTGTTATCAAGTAGCAACAATgaggcgaattc

4676   REV                        gaattcgcctCTTGACAATAAAGACTGTCTATAGCGcgacaggatcc
4677   FOR               gaattcgcctcATATAGTGAATAATATTCACCACCACcgacaggatcc
4685   REV                        ccatcctgtcgAAACACTCGTCTAGCTGTTTGTACCCgaggcgaattc
4692   FOR                        gaattcgcttcTATACAATTCCATTGTAAACAGCAACgacaggatcc
4694   FOR   gaattcgcctcACTGTTCCAAAAAGTTATGAGCAATGcgacaggatcc
```

Fig. 5. DNA sequences of oligonucleotides isolated after five rounds of selection for the binding site of GST−Sox-5 fusion protein. Uppercase letters represent the random 26 bp region of the starting 76 bp oligonucleotide population. Lowercase letters denote the common sequences flanking the random core. The consensus sequence AACAAT is shared by 12 oligonucleotides, six of which contain two such motifs. Five additional oligonucleotides bear one or two motifs of the same sequence with a 1 bp mismatch, with the difference from the consensus shown in bold type. Four other oligonucleotides contain no obvious consensus sequence (not shown). FOR and REV refer to the opposite strands of a sequence from a single plasmid clone, with the numbers of the plasmids indicated at left.

the two reduction divisions, the first haploid spermatids appear over days 19–22 (Bellve, 1979). This means that the first detection of expression of a gene can be correlated with the appearance of a particular cell-type. Sox-5 is not detected until 21 days after birth (Figure 4) and therefore Sox-5 is unlikely to be present in spermatogonia or spermatocytes and correlates with the appearance of round spermatids. This is consistent with the result of the spermatogenic cell fractionation experiment described above.

Sox-5 encodes a sequence-specific DNA binding protein

The homology of the Sry-box with known DNA binding proteins suggested that Sox-5 might also bind DNA. To test this possibility, *Sox-5* was expressed in *E.coli* as a glutathione-S-transferase (GST) fusion protein (Smith and Johnson, 1988) and used in a random oligonucleotide binding assay (Treacy *et al.*, 1991). The fusion protein binds DNA under high salt conditions (300 mM KCl), indicating strong interaction (not shown). Oligonucleotides bound to GST–Sox-5 were eluted and amplified by PCR, then used in further rounds of binding, elution and amplification, to select for sequences that bound specifically. Enrichment of specific sequence at each step was confirmed by comparison of the proportion of DNA bound by GST–Sox-5 complexed with glutathione–agarose (Chittenden *et al.*, 1991) (data not shown). After five rounds of selection, eluted DNA was amplified, cloned and sequenced. The sequences of individual clones and the derived consensus are shown in Figure 5. A sequence, AACAAT, or a minor variant, occurs in 17 of 21 sequenced clones. Six clones contain two copies of this sequence. This may be a result of Sox-5 binding to DNA cooperatively or as a dimer. Preliminary experiments suggest that the HMG-box of Sox-5 is not involved in such putative dimerisation. (F.Connor and A.Ashworth, unpublished). The Sox-5 consensus DNA binding sequence overlaps, but is distinct from the binding site of another HMG box protein, TCF-1, which recognises the sequence AACAAAG (Van de Wetering *et al.*, 1991). Similar results were obtained with a GST-fusion protein which consisted

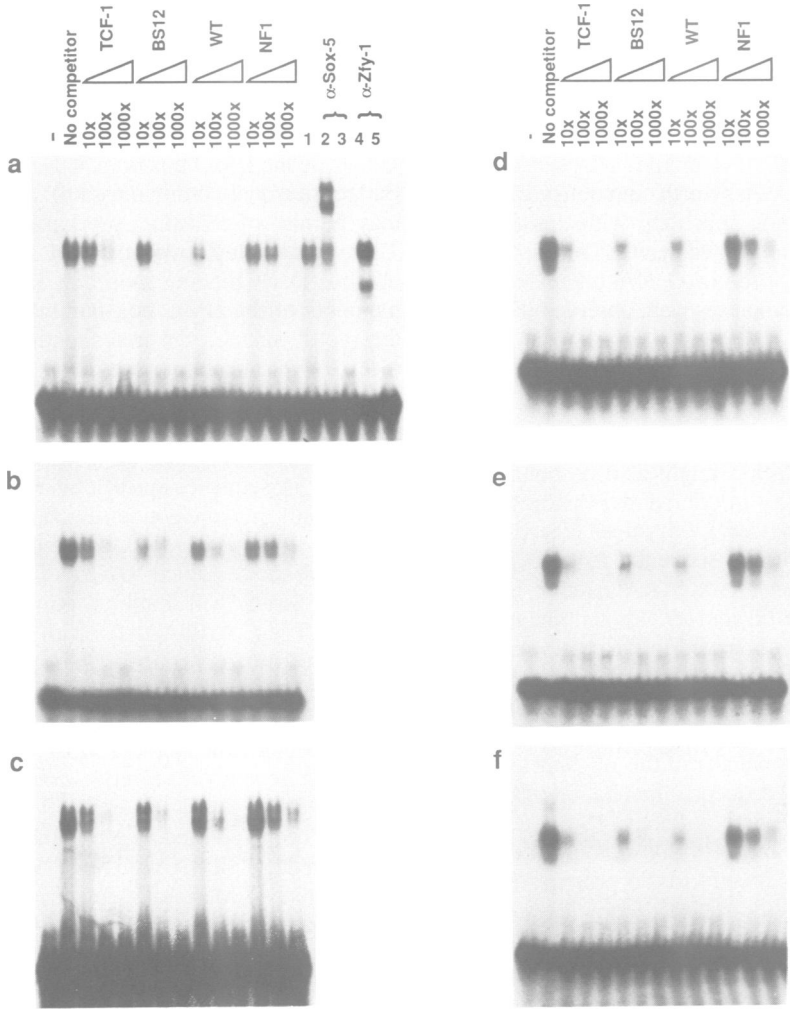

Fig. 6. Electrophoretic mobility shift assays of oligonucleotide binding by GST–Sox-5 (**a, b** and **c**) and GST–Sry (**d, e** and **f**). Unlabelled competitor oligonucleotides were included in some binding reactions at the molar excess over probe oligonucleotide indicated in the panel. Control reactions performed in the absence of DNA binding protein are indicated with a dash (–). Specific rabbit polyclonal antibodies raised against Sox-5 or against Zfy-1 (as negative control) were included in some binding reactions shown in panel a: lane 1, GST–Sox-5 alone; lane 2, GST–Sox-5 plus anti-Sox-5 antibody (137-5); lane 3, anti-Sox-5 antibody only; lane 4, GST–Sox-5 plus anti-Zfy-1 antibody (P.Taylor and A.Ashworth, unpublished); lane 5, anti-Zfy-1 antibody only. Probes used were as follows: panels a and d, Sox-5 BS12 probe; panels b and e, TCF-1 MW56 probe; panels c and f, WT12 probe. The presence of two retarded species in the GST–Sox-5 EMSAs is probably due to proteolytic clipping of the N-terminus of the GST moiety during protein preparation, as two additional slower migrating complexes are also formed in the presence of antibody 137-5, which is directed against the C-terminus of Sox-5.

of aa 149–312 of Sox-5, which contains the HMG-box and ~80 aa of flanking sequence (data not shown). This is in agreement with the results of others that the HMG-box is both necessary and sufficient for DNA binding (e.g. Jantzen *et al.*, 1990; Waterman *et al.*, 1991).

The binding of GST–Sox-5 with radiolabelled double-stranded oligonucleotides was tested using the electrophoretic mobility shift assay (EMSA) (Figure 6). The fusion protein (and unfused recombinant Sox-5, data not shown) is able to form complexes with oligonucleotides (BS12 and WT) containing the selected sequence AACAAT. The sequences of these oligonucleotides were derived from a sequence selected by Sox-5 (BS12) and from a sequence found in the promoter of the Wilm's tumour gene (*WT-1*) at −241 to −214 (Pelletier *et al.*, 1991), respectively. (Figure 6a and c). The specificity of these interactions were compared by the effect of the addition of various unlabelled oligonucleotides to the binding reaction. Complexes formed with the BS12 probe were competed with similar efficiency by BS12, WT and TCF-1 oligonucleotides, but were competed poorly by an irrelevant control oligonucleotide, NF1 (Figure 6a). Cross-competition experiments, using TCF-1 and WT oligonucleotides as probes (Figure 6b and c) confirmed that oligonucleotides containing either AACAAT or AACAAAG have similar competition characteristics. Preliminary results suggest that GST–Sox-5 or Sox-5 alone can protect a region of ~20 bp containing the AACAAT motif in a DNase I footprint assay (F.Connor and A.Ashworth, unpublished). Incubation of the DNA binding reactions with anti-Sox-5 antibody prior to EMSA demonstrated that GST–Sox-5 was contributing directly to the protein–DNA complexes as additional specific retarded complexes were observed (Figure 6a).

Recombinant Sry binds to the Sox-5 selected sequence

The similarity of Sry and Sox-5 (53% over 79 aa) suggested that sequences selected by Sox-5 might also be bound by Sry. We tested this possibility in competitive binding and EMSAs and demonstrated that GST–Sry does indeed bind specifically to the Sox-5 selected sequence AACAAT, as well as to the related sequence AACAAAG found in the TCF-1 binding site (Figure 6d and e). Recombinant Sry which lacked GST sequences behaved in the same manner (not shown). The specific binding of GST–Sry to the WT oligo (derived from the promoter region of the Wilm's tumour gene; Figure 6f) emphasizes the importance of the AACAAT sequence, as the remainder of this oligonucleotide is different to the Sox-5 BS12 sequence. Incubation of DNA binding reactions with anti-Sry antibody (S.Swift and A.Ashworth, unpublished) prior to EMSA demonstrated that GST–Sry was contributing directly to the protein–DNA complexes as additional specific retarded complexes were observed (not shown).

Discussion

The testis determining gene *SRY* and the autosomal *Sry*-like genes encode proteins that are members of the HMG-box family and are potential transcription regulating factors (Gubbay *et al.*, 1990). In order to investigate the functions and diversity of the *SRY*-related genes, we have characterised a novel member of this gene family, *Sox-5*, which is highly expressed during spermatogenesis.

Spermatogenesis is a long and complex differentiation pathway and is probably regulated by a correspondingly complex set of control mechansims. Part of this regulation is mediated at the level of transcription (Willison and Ashworth, 1987), possibly through the presence of novel transcription regulatory proteins. High levels of transcripts from the *Sox-5* gene are detected in the adult testis and using two independent approaches we have demonstrated that the 46 kDa protein is largely restricted to round spermatids, which are the haploid cells that subsequently differentiate into mature spermatozoa. Thus Sox-5 may be a cell type-specific transcription factor that could regulate the expression of spermatid-activated genes. Indeed potential Sox-5 binding sites are found in genomic regions flanking the rodent and human protamine genes (data not shown), which are highly transcribed after meiosis.

The *Sox-5* cDNA encodes a protein with hallmarks of a transcription-regulating factor, including an HMG box and regions rich in proline and acidic residues, both of which may act as transcription-activating domains (Mermod *et al.*, 1989; Jones, 1990). Comparison of the HMG-box encoded by *Sox-5* with other HMG-box proteins indicates that it is clearly a member of the *SRY* family rather than of another family, such as the TCF-1 (Castrop *et al.*, 1992) or HMG-1 families (Goodwin and Bustin, 1988). However, there appears to be little sequence homology with *Sox-1, -2* and *-3* outside the HMG box domain (J.Collignon and R.Lovell-Badge, personal communication). This suggests that *Sox-5* may be part of a distinct sub-family of *Sry*-related genes. There is a single copy of the HMG box motif and no other obvious DNA binding motifs in Sox-5. In contrast, some members of the HMG box family have multiple copies of the motif, such as the mitochondrial transcription factor, mtTF1, which has two motifs (Parisi and Clayton, 1991) and hUBF, which has six (Jantzen *et al.*, 1990; Bachvarov and Moss, 1991).

The HMG box is a rather versatile protein domain, being involved in sequence-specific and non-specific binding to DNA. Our oligonucleotide selection data shows that *Sox-5* encodes a sequence-specific HMG-box, binding preferentially to the sequence AACAAT. In the conditions of an EMSA, Sox-5 will bind to oligonucleotides containing AACAAT or AACAAAG with similar specificity, as determined by competition with unlabelled oligonucleotides. We believe that this could be explained by small differences in specificity of binding by Sox-5 to these sequences, which are not significant in the EMSA, but become significant under the repetitive, selective conditions used to identify binding sites.

It has been reported that recombinant SRY can bind to the sequences TTCAAAG and AACAAAG in DNase I protection and EMS assays, respectively (Nasrin *et al.*, 1991; Harley *et al.*, 1992). These data, taken together with the sequence similarity of Sry and Sox-5, encouraged us to test whether Sry would also bind specifically to the Sox-5 selected sequence AACAAT. This hypothesis was confirmed and may have important consequences for the potential function of this family of proteins as transcription factors. As both Sox-5 and Sry may be expressed in the same cell type in the adult testis, competitive or co-operative interactions for target sites in genes could occur. The probable overlap in the sequence specificity of Sry and related genes is reinforced by site-selection assays with GST–Sry that

result in strong enrichment of the AACAAT motif (S.Swift and A.Ashworth, unpublished). A copy of this motif is found in the promoter of the Wilm's tumour gene *Wt-1* (Pelletier *et al.*, 1991), but it remains to be determined whether this reflects any functional interaction with Sry or related genes.

Preliminary experiments suggest that Sox-5 is unable to activate transcription from a minimal promoter linked to single or multimerized Sox-5 binding sites (data not shown). This raises the intriguing possibility that Sox-5, like LEF-1 (Travis *et al.*, 1991), can only activate transcription from binding sites within the context of an appropriate promoter. The recent demonstration that HMG-box proteins such as LEF-1 and Sry can induce significant bending in DNA upon binding (Giese *et al.*, 1992) suggests that transcription activation by these factors might be mediated by the promotion of protein–protein interaction. Experiments are underway to determine whether Sox-5 can bend DNA and to characterize *in vivo* binding sites, which should lead to an understanding of the role of this protein in transcriptional regulation in the testis.

Materials and methods

Reverse transcription and PCR

Poly(A)$^+$ RNA was isolated from various mouse (male Parkes outbred) tissues using a Fast-Track kit (Invitrogen) and reverse transcribed as described (Denny and Ashworth, 1991). Reactions were incubated at 37°C for 60 min. Degenerate oligonucleotide primers corresponding to all of the potential codons which encode aa residues 17–23 (sense strand) and 78–84 (antisense strand) in *SRY* and *SRY*-related genes (Gubbay *et al.*, 1990) were used in PCR reactions. Sequences of these primers are shown in Figure 1b. PCRs were performed in 50 μl reaction volumes containing 2.5 units of Taq polymerase (Perkin–Elmer Cetus), in a buffer consisting of 60 mM KCl, 15 mM Tris–HCl, pH 8.8 (at 20°C), 2.25 mM $MgCl_2$, each dNTP at 0.25 mM and 10 pmol of each primer. Approximately 25 ng of cDNA or 1 μl of phage supernatant from an amplified cDNA library was added to the reaction, which was then subjected to 35–40 cycles of PCR in a Techne PHC-1 programmable thermal cycler, each cycle being 30 s incubation at 94°C, 120 s at 37°C, 60 s at 72°C, with a 5 min final extension at 72°C. The products of the reaction were purified by gel electrophoresis and ligated into the *Eco*RV site of pBluescript (Stratagene) using standard techniques (Sambrook *et al.*, 1989).

Screening cDNA libraries

We screened an amplified, adult mouse testis cDNA library in λZAPII XR (Stratagene) derived from ~2 × 10^6 primary recombinants (A.S.Goldsborough, P.Denny and A.Ashworth, manuscript in preparation) using a PCR-based strategy. The library consists of several independent fractions, each corresponding to specific pools amplified from ~40 000 recombinant phage. Individual library fractions were screened by PCR with the *SRY*-box primers (above) and those which were positive in this assay were then screened by hybridization, according to standard procedures (Sambrook *et al.*, 1989). Hybridization probes consisted of an equimolar mixture of the *Sry*-box cDNA inserts from *Sox-5* and *-6* (Denny *et al.*, 1992). Recombinant pBluescript II DNAs were excised from λZAPII phage as recommended by the manufacturer (Stratagene). Plasmids pD578, pD597 and pD598 were obtained in the initial screening and the longest clone, pD588, was identified by PCR, using a combination of *Sox-5*-specific and vector -specific primers (not shown).

Sequencing and gene nomenclature

Sequencing was performed on both strands of double stranded templates using T7 DNA polymerase (Pharmacia) and conditions described in the Sequenase kit (USB). It has been proposed (R.Lovell-Badge, personal communication) that the *SRY*-related genes that share the HMG-box motif be named the *SOX*-genes for 'SRY-box'. In accordance with standardized genetic nomenclature, lower case italic symbols are used for murine genes and upper case italic for the human. When referring to the gene family as a collective group, upper case italic is used, i.e. *SOX*.

Expression of Sox-5 and Sry as glutathione-S-transferase (GST) fusion proteins in bacteria

Nucleotides 144–1415 from the *Sox-5* cDNA pD588, encoding aa 16–392, were subcloned into the pGEX-3X expression vector (Smith and Johnson, 1988). Nucleotides 98–370 of mouse *Sry*, encompassing the HMG box (Gubbay *et al.*, 1990), were amplified by PCR from genomic DNA, cloned into pGEX-2T and sequenced. GST-fusion proteins were expressed in BL21 (DE3) pLysS host bacteria (Studier, 1991) and purified using glutathione–agarose beads as described (Smith and Johnson, 1988). Proteins were >90% pure as determined by Coomassie blue staining of SDS–polyacrylamide gels.

Preparation of antisera, immunoblotting and COS cell transfection

A 15 residue peptide consisting of the predicted C-terminal 14 aa of Sox-5, together with an N-terminal cysteine residue was synthesized and crosslinked to keyhole limpet haemocyanin (KLH) as described (Sambrook *et al.*, 1989). Antibodies were raised in rabbits immunised (Harlow and Lane, 1988) with the peptide–KLH conjugate and were purified by affinity chromatography on peptide-Reactigel (Pierce) columns, as recommended by the manufacturer. Protein extracts from whole tissues or cultured COS cells were prepared by homogenization in ice-cold Tris-buffered saline, then adjusting the homogenate to 3% (w/v) SDS, 5% β-mercaptoethanol (v/v), 10% (v/v) glycerol, 0.125 M Tris–HCl, pH 6.8 and heating to 100°C for 5 min. Viscosity was reduced by brief sonication. SDS–PAGE of total protein extracts was performed on 10% polyacrylamide gels and proteins transferred to nitrocellulose filters (Towbin *et al.*, 1970). Filters were blocked in a solution of 1 M glycine, 5% (w/v) non-fat milk powder, 1% (w/v) albumen–5% (v/v) newborn calf serum and incubated with affinity purified rabbit antibody, diluted in the same solution. The primary antibody was detected using anti-rabbit IgG–horseradish peroxidase conjugate (Promega) and enhanced chemiluminescence (ECL) reagents (Amersham). Male mouse germ cells were fractionated and protein extracts prepared as described (Willison *et al.*, 1990). Testes from four adult mice were dissected and a suspension of cells produced by treatment with collagenase and trypsin, as described (Romrell *et al.*, 1976). Krebs–Ringer bicarbonate (EKRB) medium (Romrell *et al.*, 1976) supplemented with 11 mM glucose, 1 mM glutamine, 1 mM pyruvate and 6 mM lactate, was used throughout the preparation. Cells in 100 ml of EKRB plus 0.5% BSA were layered on a 850 ml density gradient of 2–4% BSA, in a CelSep chamber (DuPont) with a 50 ml cushion of 10% BSA at the bottom and sedimented under unit gravity for 90 min at room temperature. Seventeen fractions of 50 ml were collected from the bottom of the gradient (densest region; the first fraction, containing the cushion, was discarded) and cell purity and numbers analysed by light microscopy. Immunoblots of protein from each fraction were analysed as described above and following removal of antibodies (Amersham ECL), filters were probed with a monoclonal antibody to TCP-1 (monoclonal 91a; Willison *et al.*, 1989) as a control.

Sox-5 cDNA clone pd588, encoding the full-length protein, was subcloned into eukaryotic expression vector pcDNA I (Invitrogen), producing plasmid pd740. This plasmid, or as a negative control, pcDNA I, was transfected into COS cells by electroporation. COS cells were harvested, washed and electroporated using a Gene-Pulser with capacitance extender (BioRad) as described (Chu *et al.*, 1987). Cells were harvested using trypsin–versene, washed once with medium–10% fetal calf serum, washed twice with 20 mM HEPES pH 7.05, 137 mM NaCl, 5 mM KCl, 0.7 mM Na_2HPO4, 6 mM dextrose and electroporated with 10 μg plasmid DNA and 10 μg carrier salmon sperm DNA at 250 V/125 μF. Cells were re-plated 10 min after electroporation and cultured for 72 h before protein extracts were prepared as described above.

Identification of DNA binding site of Sox-5

Sequences binding to GST–Sox-5 were selected from a pool of double-stranded, 76mer oligonucleotides containing a central stretch of 26 random base-pairs, flanked by two unique sequences of 25 bp each, which allow recovery of minute quantities of bound DNA by PCR amplification (Pollock and Treisman, 1990). The sequences of these oligonucleotides were: Random sequence 76mer oligonucleotide (R76), 5′-CAGGTCAGATCAGCGGATCCTGTCG$(N)_{26}$GAGGCGAATTCAGTGCATGTGCAGC-3′, 'Forward' primer, 5′-GCTGCACATGCACTGAATTCGCCTC; 'Back' primer, 5′-CAGGTCAGATCAGCGGATCCTGTCG. Random oligonucleotide binding assays were performed essentially as described (Treacy *et al.*, 1991). Briefly, 1 μg of purified fusion protein was electrophoresed on a 10% SDS–PAGE gel, transferred to a nitrocellulose filter and the filter incubated in buffer A (20 mM HEPES–NaOH, pH 7.9, 1 mM DTT, 10% (v/v) glycerol, 0.01% (v/v) NP-40) containing 50 mM KCl and 5% (w/v) low-fat milk powder for 4 h at 4°C. The ^{32}P-labelled pool of random oligonucleo-

tides (10^7-10^8 c.p.m./μg) was prepared by primed synthesis using the forward primer annealed to the random 76mer oligonucleotide template (Pollock and Treisman, 1990), then added to the filter and incubated overnight at 4°C. Washes were performed at 4°C in buffer A adjusted to 100 mM KCl, (2 × 10 min) and then in buffer A adjusted to 300 mM KCl (1 × 10 min). Filters were autoradiographed, the area of filter corresponding to bound DNA excised and the DNA eluted by heating to 100°C in H_2O. Eluted DNA was amplified by 30 cycles of PCR using 'forward' and 'back' primers, and the 76 bp product purified on 3% NuSieve−1% agarose gels. For subsequent rounds of binding, the 76 bp product was labelled by 18 cycles of PCR in the presence of ^{32}P-labelled nucleotide, as described (Pollock and Treisman, 1990). Sequence enrichment was tested by comparing the efficiency with which labelled DNA from each round of selection and amplification bound to the GST−Sox-5 fusion protein complexed with glutathione−agarose beads, as described (Chittenden *et al.*, 1991). DNA from the fourth round of selection which remained bound to GST−Sox-5 after washing in buffer A + 300 mM KCl was eluted and amplified as described above, cloned into pBluescript, the inserts from 21 clones amplified by PCR and the purified PCR products sequenced (A.Ashworth and A.S.Goldsborough, unpublished).

Electrophoretic mobility shift assays

Electrophoretic mobility-shift assays were performed as described by Van de Wetering *et al.* (1991). Briefly, 10 ng of purified GST-fusion protein was pre-incubated in 20 μl of binding buffer (10 mM HEPES (pH 7.9), 60 mM KCl, 1 mM DTT, 1 mM EDTA, 0.25 mg/ml bovine serum albumin and 12% glycerol) with appropriate competitor DNA for 5 min at room temperature, then 10 fmol (~20 000 c.p.m.) of ^{32}P-labelled, annealed oligonucleotides was added and incubated for a further 20 min at room temperature. Where indicated, 1 μl of antibody was added to the pre-incubation. The reactions were electrophoresed on a non-denaturing 4% acrylamide gel in 0.5 × TBE buffer at room temperature at 10 V/cm for 90 min. Gels were dried prior to autoradiography. Probes and competitors ('top' strand of double-stranded oligonucleotides): Sox-5 BS12 5′-tcgaGCACTAAAACAATTCAAGCCCGGGG; WT 5′-gggAATTATC-TATCCCGAACAATTTCACCTT (Pelletier *et al.*, 1991); TCF (MW56)5′-GGGAGACTGAGAACAAAGCGCTCTCACAC (Van de Wetering *et al.*, 1991); NF1 5′-cgcatTTGGCAGCTTGCCAAGGATCCTTGGCAGCT-TGCCAAG (R.Nicolas and G.Goodwin, unpublished). Lower case characters indicate residues added to allow fill-in labelling by Klenow and/or ligation of oligonucleotides. Underlining indicates the sequences bound by Sox-5 and/or Sry. Probes were labelled by Klenow DNA polymerase I with [^{32}P]dCTP and unlabelled dNTPs, where appropriate.

Acknowledgements

We would like to thank A.S.Goldsborough, G.Hynes, R.Nicolas, P.Taylor and K.Weston for gifts of reagents, H.King for oligonucleotide synthesis and J.Westwood for peptide synthesis. We are also grateful to M.Crompton, G.Goodwin, G.Lang, R.Nicolas, K.Orchard, P.Taylor, M.N.Treacy and K.Willison for helpful discussions. We would also like to thank J.Collignon and R.Lovell-Badge for sharing data prior to publication. This work was supported by the Cancer Research Campaign, the Medical Research Council and by a Wellcome Trust Prize Studentship to F.C.

References

Bachvarov,D. and Moss,T. (1991) *Nucleic Acids Res.*, **19**, 2331−2335.
Bellve,A.R. (1979) In Finn,C.A. (ed.), *Oxford Reviews of Reproductive Biology*. Oxford University Press, Oxford, Vol. 1, pp. 159−261.
Burgoyne,P.S., Buehr,M., Koopman,P., Rossant,J. and McLaren,A. (1988) *Development*, **102**, 443−450.
Castrop,J., Van Norren,K. and Clevers,H. (1992) *Nucleic Acids Res.*, **20**, 611.
Chittenden,T., Livingston,D.M. and Kaelin,W.G.,Jr (1991) *Cell*, **65**, 1073−1082.
Chu,G., Hayakawa,H. and Berg,P. (1987) *Nucleic Acids Res.*, **15**, 1311−1326.
Denny,P. and Ashworth,A. (1991) *Gene*, **106**, 221−227.
Denny,P., Swift,S., Brand,N., Dabhade,N., Barton,P. and Ashworth,A. (1992) *Nucleic Acids Res.*, **20**, 2887.
Garnier,J., Osguthorpe,D.J. and Robson,R. (1978) *J. Mol. Biol.*, **120**, 97−120.
Giese,K., Cox,J. and Grosschedl,R. (1992) *Cell*, **69**, 185−195.
Goodwin,G. and Bustin,M. (1988) In Kahl,G. (ed.), *Architecture of eukaryotic genes*. VCH Verlagsgesellschaft mbH, Weinheim, pp. 187−205.
Gubbay,J., Collignon,J., Koopman,P., Capel,B., Economou,A., Munsterberg,A., Vivian,N., Goodfellow,P. and Lovell-Badge,R. (1990) *Nature*, **346**, 245−250.
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**, 453−456.
Harlow,E. and Lane,D. (1988) *Antibodies: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Jantzen.H.-M., Admon,A., Bell,S.P. and Tjian,R. (1990) *Nature*, **344**, 830−836.
Jones,N. (1990) *Semin. Cancer Biol.*, **1**, 5−17.
Kelly,M., Burke,J., Smith,M., Klar,A. and Beach,D. (1988) *EMBO J.*, **7**, 1537−1547.
Koopman,P., Munsterberg,A., Capel,B., Vivian,N. and Lovell-Badge,R. (1990) *Nature*, **348**, 450−452.
Koopman,P., Gubbay,J., Vivian,N., Goodfellow,P. and Lovell-Badge,R. (1991) *Nature*, **351**, 117−121.
McLaren,A. (1988) *Trends Genet.*, **4**, 153−157.
Mermod,N., O'Neill,E.A., Kelly,T.J. and Tjian,R. (1989) *Cell*, **58**, 741−753.
Nasrin,N., Buggs,C., Kong,X.F., Carnazza,J., Goebl,M. and Alexander-Bridges,M. (1991) *Nature*, **354**, 317−320.
Parisi,M.A. and Clayton,D.A. (1991) *Science*, **252**, 965−969.
Pelletier,J., Schalling,M., Buckler,A.J., Rogers,A., Habre,D.A. and Housman,D. (1991) *Genes and Dev.*, **5**, 1345−1356.
Pollock,R. and Treisman,R. (1990) *Nucleic Acids Res.*, **18**, 6197−6204.
Romrell,L.J., Bellve,A.R. and Fawcett,D.W. (1976) *Dev. Biol.*, **49**, 119−131.
Rubin,M.R., Toth,L.E., Patel,M.D., d'Eustachio,P. and Chi Nguyen-Huu,M. (1986) *Science*, **233**, 663−667.
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.
Shackleford,G.M. and Varmus,H.E. (1987) *Cell*, **50**, 89−95.
Sinclair,A.H. *et al.* (1990) *Nature*, **346**, 240−244.
Smith,D.B. and Johnson,K.S. (1988) *Gene*, **67**, 31−40.
Studier,F.W. (1991) *J. Mol. Biol.*, **219**, 37−44.
Towbin,H., Staehelin,T. and Gordon,J. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 4350−4354.
Treacy,M.N., He,X. and Rosenfeld,M.G. (1991) *Nature*, **350**, 577−584.
Travis,A., Amsterdam,A., Belanger,C. and Grosschedl,R. (1991) *Genes Dev.*, **5**, 880−894.
Van de Wetering,M., Oosterwegel,M., Dooijes,D. and Clevers,H. (1991) *EMBO J.*, **10**, 123−132.
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,G.R. (1989) *Nucleic Acids Res.*, **17**, 1197−1214.
Willison,K. and Ashworth,A. (1987) *Trends Genet.*, **3**, 351−355.
Willison,K., Lewis,V., Zuckerman,K.S., Cordell,J., Dean,C., Miller,K., Lyon,M.F. and Marsh,M. (1989) *Cell*, **57**, 621−632.
Willison,K.R., Hynes,G., Davies,P., Goldsborough,A. and Lewis,V.A. (1990) *Genet. Res.*, **56**, 193−201.
Wolgemuth,D.J., Viviano,C.M., Gizang-Ginsberg,E., Frohman,M.A., Joyner,A.L. and Martin,G.R. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 5813−5817.

Received on April 2, 1992; revised on June 23, 1992