

University of Nebraska - Lincoln  
DigitalCommons@University of Nebraska - Lincoln

---

Virology Papers

Virology, Nebraska Center for

---

2000

# Isolation, Characterization, and Differential Expression of the Murine Sox-2 Promoter

Matthew S. Wiebe

*University of Nebraska-Lincoln, [mwiebe2@unl.edu](mailto:mwiebe2@unl.edu)*

Phillip J. Wilder

*Eppley Institute for Research in Cancer and Allied Diseases at the University of Nebraska Medical Center, USA*

David Kelly

*Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center*

Angie Rizzino

*Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, [arizzino@unmc.edu](mailto:arizzino@unmc.edu)*

Follow this and additional works at: <http://digitalcommons.unl.edu/virologypub>

---

Wiebe, Matthew S.; Wilder, Phillip J.; Kelly, David; and Rizzino, Angie, "Isolation, Characterization, and Differential Expression of the Murine Sox-2 Promoter" (2000). *Virology Papers*. 285.

<http://digitalcommons.unl.edu/virologypub/285>

This Article is brought to you for free and open access by the Virology, Nebraska Center for at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Virology Papers by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Published in *Gene* 246:1–2 (2000), pp. 383–393.

Copyright © 2000 Elsevier Science B.V. Used by permission.

Submitted November 16, 1999; revised February 1, 2000; accepted February 9, 2000.

# Isolation, Characterization, and Differential Expression of the Murine Sox-2 Promoter

Matthew S. Wiebe,<sup>1,2</sup> Philip J. Wilder,<sup>1</sup> David Kelly,<sup>1</sup> and Angie Rizzino<sup>1,2</sup>

1. Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, Omaha, NE 68198-6805, USA
2. Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, NE 68198-6805, USA

*Corresponding author* – Angie Rizzino, Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, Omaha, NE 68198-6805, USA; phone 402-559-6338, email [arizzino@unmc.edu](mailto:arizzino@unmc.edu)

## Abstract

Sox proteins are expressed at many stages of development and in numerous tissues. The transcription factor Sox-2 is first expressed throughout the inner cell mass and subsequently becomes localized to the primitive ectoderm, developing central nervous system, and the lens. Sox-2 is also highly expressed in F9 embryonal carcinoma cells but becomes undetectable following differentiation of these cells. In this study, we have isolated, sequenced, and performed the first characterization of the Sox-2 promoter of any species. Approximately 2 kb of the Sox-2 5'-flanking region has been sequenced and the primary transcription start site mapped by primer extension analysis. Additionally, two positive regulatory regions within the promoter region have been identified. We also show that expression of Sox-2 promoter/reporter gene constructs is reduced in differentiated EC cells as compared to their undifferentiated counterparts. Furthermore, we have identified a consensus inverted CCAAT box motif present in the Sox-2 promoter. Mutagenesis of this site significantly reduces the expression of Sox-2 promoter/reporter constructs. We also demonstrate that this CCAAT box motif can bind the trimeric transcription factor NF-Y.

**Keywords:** embryonal carcinoma cells, FGF-4, NF-Y, SRY, UTF1

**Abbreviations:**  $\beta$ Gal,  $\beta$ -galactosidase; *cat*, chloramphenicol acetyl-transferase gene; *CYP1A1*, cytochrome P4501A1 gene; DMEM, Dulbecco's modified Eagle's medium; dsODN, double-stranded oligo-nucleotide; EC, embryonal carcinoma; FBS, fetal bovine serum; *FGF-4*, fibroblast growth factor

4 gene; HMG, high mobility group; *TβR-II*, type II transforming growth factor-β receptor gene; *UTF1*, undifferentiated embryonic cell transcription factor 1 gene; *UTR*, untranslated region.

## 1. Introduction

Members of the Sox family were originally grouped together on the basis of their high homology to the HMG (high mobility group) domain of the testis-determining gene *SRY*. The HMG domain of Sox proteins is capable of binding in the minor groove of DNA in a sequence-specific manner (van de Wetering et al., 1993). In addition to the presence of a DNA binding domain, some Sox factors have been shown to contain modular transactivation domains (van de Wetering et al., 1993; Kuhlbrodt et al., 1998a, b; Nowling et al., 2000). These Sox factors are expressed at myriad stages during development and in numerous tissue types (Wegner, 1999) and have been found to act as transcriptional activators for a variety of genes that are developmentally regulated (van de Wetering et al., 1993; Yuan et al., 1995). Hence, the Sox protein family is likely to play pivotal roles in the control of development.

The transcription factor Sox-2 exhibits many of the hallmark features of the Sox family. Sox-2 expression is regulated both temporally and spatially. It is first expressed throughout the inner cell mass and subsequently becomes localized primarily to the primitive ectoderm and the developing nervous system (Collignon et al., 1996). It is also expressed in the lens later in development (Kamachi et al., 1995). Genes such as fibroblast growth factor 4 (*FGF-4*) and undifferentiated embryonic cell transcription factor 1 (*UTF1*), which are expressed early in development, rely on enhancer elements that contain essential binding sites for Sox-2 (Yuan et al., 1995; Johnson et al., 1998; Nishimoto et al., 1999). In addition to acting as a transcriptional activator, Sox-2 is able to act as a repressor of transcription in some contexts, as observed in the regulation of the osteopontin gene (Botquin et al., 1998). Furthermore, Pevny et al. (1998) refer to unpublished reports that claim that Sox-2 homozygous null embryos die soon after implantation, reinforcing the argument that Sox-2 is essential during embryogenesis.

Despite the mounting evidence that Sox-2 is a key player in early developmental gene regulation, little is known about the transcriptional regulation of the *Sox-2* gene itself. Embryonal carcinoma (EC) cells provide an excellent cell model system in which to study the impact of cell differentiation on the expression of genes regulated during embryogenesis (Rizzino, 1989). *Sox-2* is highly expressed in undifferentiated EC cells, but Sox-2 mRNA becomes undetectable upon differentiation of EC cells (Yuan et al., 1995; Johnson et al., 1998), which suggests that *Sox-2* is transcriptionally downregulated in this system. In this report, we address three main questions in an effort to elucidate the mechanisms by which *Sox-2* is differentially regulated in EC cells. First, are positive regulatory elements present in the 5'-flanking region of the *Sox-2* gene? To answer this question, we describe the cloning and sequencing of the *Sox-2* promoter, the mapping of the primary transcription start site of the *Sox-2* gene, and the initial characterization of the *Sox-2* promoter using promoter/reporter gene constructs. We demonstrate that the region of the *Sox-2* gene between -528 and +238 contains regulatory elements capable of activating gene transcription. Second, are regulatory elements present in the promoter subject to differential regulation when EC

cells differentiate? We demonstrate that reporter constructs driven by regions of the *Sox-2* promoter are actively expressed when transfected into undifferentiated EC cells, but their expression is diminished significantly when the same constructs are transfected into EC-differentiated cells. Finally, we have identified a putative CCAAT box motif present in the *Sox-2* promoter. Mutagenesis of this CCAAT box in promoter/reporter gene constructs reduced the ability of the *Sox-2* promoter to activate transcription compared to the wild-type promoter, indicating that the CCAAT box motif is important for *Sox-2* expression. Furthermore, gel mobility supershift analysis revealed that the trimeric transcription factor NF-Y is capable of binding the *Sox-2* CCAAT box motif.

## 2. Materials and methods

### 2.1. Isolation and subcloning of murine *Sox-2* genomic DNA

A bacteriophage P1-based mouse strain 129/Ola ES cell genomic library was screened by PCR using primers specific for the murine *Sox-2* 5'-untranslated region (*UTR*): 5'-TCA GAGTTGTCAAGGCAGAGA-3' (sense) and 5'-CGCCGCGATTGTTGTGATTAGT-3' (antisense). These primers yield a 166 bp product from both genomic DNA and the *Sox-2* P1 genomic clone, which was obtained from Genome Systems (St. Louis, Missouri). Bacteria containing the P1 clone were grown, and we isolated DNA using a kb-100 (Magnum) kit (Genome Systems). Subcloning of the P1 clone was accomplished using standard techniques. Briefly, the P1 clone was digested with various restriction enzymes, the resulting fragments purified and ligated into the multiple cloning site of pBluescript II KS+ and transformed into competent bacteria. Standard blue/white screening methods and PCR were used to identify bacterial colonies containing the desired insert.

In the present study, two separate clones were selected for detailed characterization. They are referred to as Sox-2-5'Cl and Sox-2-3'Cl, and each contains one end of the *Sox-2* gene. The PCR primers mentioned above, which amplify the sequence from +17 to +183 near the beginning of the 5'-*UTR*, were used to screen for inserts containing the 5'-flanking region; this identified Sox-2-5'Cl. A second set of primers was used to screen for inserts containing the remainder of the *Sox-2* gene, including the 3' flanking region of the gene. The primers 5'-CGCATGTATAACATGATGGAGACG-3' (sense) and 5'-CATGGGCCTCTTGACGCGG-3' (antisense), amplify the region of the *Sox-2* gene between +404 and +547. In this manner, clone Sox-2-3'Cl was identified as containing the 3' end of the *Sox-2* gene. The ends of the cloned inserts were sequenced to determine orientation, and restriction enzyme sites were mapped. Sox-2-5'Cl was the product of a *Bgl*II digestion and includes almost 6 kb of genomic sequence, while Sox-2-3'Cl comprises an *Xho*I genomic fragment of about 10 kb (fig. 1A). These two clones overlap by 40 bp within the 5'-*UTR* of the *Sox-2* gene.

### 2.2. DNA sequencing and analysis

DNA sequencing was performed in the Molecular Biology Core Facility at the Eppley Institute using an ABI 373 DNA Sequencer (Perkin-Elmer, Norwalk, Connecticut). The regions flanking the *Sox-2* transcript and the region containing a 35 bp discrepancy relative to published cDNA sequences, which are reported in this study, were sequenced on both

strands. Numbering of the Sox-2 promoter sequence has been determined by setting the 5'-nucleotide of the Sox-2 cDNA sequence reported by Collignon et al. (1996) as +1. Alignment of sequences was accomplished using version 1.1 of the Sequence Navigator program (PE/ABI). The TFSEARCH Internet site (<http://pdap1.trc.rwcp.or.jp/research/db/TFSEARCH.html>), the BLAST network service at the National Center for Biotechnology (NCBI), and Genetics Computer Group (GCG) programs (Program Manual for the Wisconsin Package, Genetics Computer Group) were used in the analysis of the DNA sequence.

### 2.3. Southern blot analysis

Southern blot analysis was performed using 10 µg aliquots of genomic DNA (from R1 ES cells) digested with *EcoRI*, *EcoRV*, or *HindIII* and transferred onto a nylon membrane. The blot was probed with a 1 kb fragment from a Sox-2 cDNA clone, pBS-Sox-2 (Johnson et al., 1998). The fragment was isolated using a *HindIII* site within the Sox-2 coding region and a downstream *HindIII* site within the multiple cloning site of the cloning vector, and then labeled with <sup>32</sup>P-dCTP using the Rediprime II labeling kit (Amersham, Arlington Heights, Illinois). The most stringent wash employed 0.1 × SSPE buffer at 60°C (1 × SSPE consists of 150 mM NaCl, 10 mM NaH<sub>2</sub>PO<sub>4</sub> and 1 mM EDTA at pH 7.4). The blot was then exposed to a phosphorimager cassette and visualized using a PhosphorImager (Molecular Dynamics).

### 2.4. Primer extension analysis

The transcriptional start site utilized by the Sox-2 promoter was mapped using the anti-sense primer Sox2PE (5'-CTGGCGGAGAATAGTTGGGGGAAGCGGAG-3'), which is complementary to base pairs +239 to +269 of the Sox-2 cDNA sequence (Collignon et al., 1996). This primer was end-labeled with [ $\gamma$ -<sup>32</sup>P]ATP using T4 Polynucleotide Kinase (New England Biolabs, Beverly, Massachusetts). Approximately 10<sup>5</sup>cpm of labeled primer was incubated with 3 µg poly(A)<sup>+</sup> RNA isolated from undifferentiated F9 EC cells using a FastTrack 2.0 mRNA Isolation Kit (Invitrogen, San Diego, California) in a reaction brought to a total volume of 30 µl with hybridization buffer (40 mM PIPES pH 6.4, 1 mM EDTA, 0.4 M NaCl, and 80% formamide). The reaction mix was denatured at 85°C for 10 min and allowed to hybridize overnight at 30°C. The RNA-primer duplexes were then ethanol precipitated and resuspended in 24 µl of a reverse transcription master mix (50 mM Tris-Cl pH 8.3, 75 mM KCl, 3 mM MgCl<sub>2</sub>, 10 mM dithiothreitol, 1 mM each dNTP, 40 units Rnasin (Promega, Madison, Wisconsin), and 50 µg/ml actinomycin D). Primer extension was initiated by adding 200 units Moloney murine leukemia virus reverse transcriptase (Promega, Madison, Wisconsin) and incubated at 42°C for 1 h. Thereafter, 2 µl of RNase ONE (Promega, Madison, Wisconsin) was added and the reaction incubated at 37°C for 30 min. The reaction was brought to 450 µl with dH<sub>2</sub>O and concentrated to approx. 5 µl using a MicroCon YM-10 Centrifugal Filter Device (Millipore, Bedford, Massachusetts). The extension reaction was electrophoresed on a 6% polyacrylamide-7 M urea gel, and visualized by Phosphorimager (Molecular Dynamics, Sunnyvale, California). The length of the cDNA extension product was determined by comparison to a radiolabeled pBR322-*MspI* digest (New England Biolabs, Beverly, Massachusetts) electrophoresed in the adjacent lane.

### 2.5. Cell culture and transient transfection

F9 EC cells were maintained in Dulbecco's modified Eagle's medium (DMEM) (Life Technologies) supplemented with 10% fetal bovine serum (FBS) (Hyclone). F9 EC cells were seeded at  $5 \times 10^5$  cells per 100 mm dish 24 h prior to transfection. Cells were transfected by the calcium phosphate precipitation method as described previously (Ma et al., 1992), incubated with the precipitate for 14–16 h, and then refed with DMEM containing 10% FBS. Cells were harvested the following day and cell extracts were prepared. Chloramphenicol acetyl-transferase (CAT) and  $\beta$ -galactosidase ( $\beta$ Gal) activities were determined as reported previously (Johnson et al., 1998). The total amount of transfected DNA was kept constant by addition of null vectors. To adjust for transfection efficiency, all transfections were normalized with pCH110 (SV40  $\beta$ Gal) (Pharmacia, Piscataway, New Jersey). All transfections were performed in duplicate plates and repeated three times with representative transfections shown. Error bars shown represent 1 S.D. and were computer generated. Plasmid DNA was purified using Qiagen tip-500 columns.

For transfection of EC-differentiated cells, F9 EC cells were seeded at  $1 \times 10^5$  cells per 100 mm dish in DMEM containing 10% FBS and 5  $\mu$ M retinoic acid (RA). After 72 h, the cells were transfected as described above and refed 14–16 h post-transfection with DMEM containing 10% FBS and 5  $\mu$ M RA. Cells were harvested 72 h later, cell extracts prepared, and CAT and  $\beta$ Gal activities determined as described above.

### 2.6. Promoter-reporter constructs and site-directed mutagenesis

The following PCR primers were used in the design of the *Sox-2* promoter/reporter gene constructs: 5'-GCTAAGCTTCTGCAGGACAAGCTTGGAAAGAGGGGC-3' (528-Sense), 5'-GCCCTCGAGCTCTTCTCTGCCTTGACAACCTCC-3' (41-Antisense), and 5'-GCGCTCGA GCAATTGGGATGAAAAACAGGC-3' (238-Antisense). The underlined primer sequence indicates either the presence of a *Hind*III or *Xho*I site included for screening of positive clones.

To generate *Sox-2*<sup>-528/+238</sup>, PCR was performed using *Sox-2*-5'CI, which contains the 5'-flanking region of the *Sox-2* genomic sequence, as a template with the primer pair of 528-Sense and 238-Antisense. The resulting product was cloned into the *Hind*III and *Xho*I sites of the *cat* expression vector pBLCAT7 (Rosfjord et al., 1994). *Sox-2*<sup>-528/+41</sup> was generated in the same fashion using the primer pair of 528-Sense and 41-Antisense. To obtain *Sox-2*<sup>-72/+41</sup>, *Sox-2*<sup>-528/+41</sup> was digested with the restriction enzymes *Hind*III and *Nae*I, resulting in a 5.3 kb and 460 bp fragment. The 5.3 kb fragment, which lacks the region of the *Sox-2* promoter between -528 and -72, was isolated by agarose gel electrophoresis, blunt-ended by treatment with Klenow DNA Polymerase, and recircularized by ligation. The *Sox-2*<sup>-72/+238</sup> construct was generated in the same manner by digesting the *Sox-2*<sup>-528/+238</sup> construct with *Hind*III/*Nae*I. The identity of the resulting four *Sox-2* promoter constructs was confirmed by sequencing.

Mutagenesis of the inverted CCAAT box sequence in the *Sox-2*<sup>-528/+238</sup> construct was performed using the QuikChange (Stratagene, La Jolla, California) method, using the following PCR primer pair: CMut-Sense: 5'-AGCCGCGCTCGCTGCAGCTGTATCGGAAACC CATTTC-3', CMut-Antisense: 5'-GAATAAATGGGTTTCCGATACAGCTGCAGCTA GCGCCG-3'.

These primers are complementary to each other as well to the region of the template surrounding the inverted CCAAT box; however, the underlined primer sequence contains a *PvuII* site and scrambles the CCAAT box while leaving the flanking sequence unchanged. To mutagenize the CCAAT box in the Sox-2<sup>-528/+238</sup> construct, a PCR reaction was performed with the CMut-Sense and CMut-Antisense primer pair; however, only the CMut-Sense primer was present for the first five cycles in order to decrease the chance of primer-dimer formation. The PCR reaction was subjected to *DpnI* digestion to eliminate the parent vector, and transformed into DH5 $\alpha$ F' competent cells. A positive clone identified by *PvuII* digestion was verified by sequencing and designated Sox-2Mut<sup>-528/+238</sup>.

### 2.7. Nuclear extract preparation and electrophoretic gel mobility shift assay

Nuclear proteins were extracted from undifferentiated F9 EC cells using the method of Dignam et al. (1983) as modified in this laboratory (Scholtz et al., 1996). All extracts were dialyzed and stored (-80°C) in a buffer consisting of 20 mM HEPES (pH 7.9), 10% glycerol, 100 mM KCl, 0.5 mM dithiothreitol, and protease and phosphatase inhibitors as described previously (Scholtz et al., 1996).

Gel mobility shift analyses were based on the method of Fried and Crothers (1981) as modified by this laboratory (Kelly et al., 1995). Complementary oligonucleotides were annealed for each probe or competitor, and the resulting double-stranded oligodeoxynucleotide (dsODN) probes were labeled with  $\alpha$ [<sup>32</sup>P]dCTP by Klenow fill-in reaction of the single-stranded regions at the end of each ODN (lower-case bases). The dsODN probe includes the CCAAT box motif (bold) and flanking sequences from -77 to -37. The sequence of the wild-type Sox-2 CCAAT dsODN probe (sense strand) is: 5'-agcCGGCGCTCGCTG **ATTGGCCGCGGAAACCCATTTATTC**-3'. The mutant CCAAT dsODN probe sequence is 5'-agcCGGCGCTCGCTGCAGCTGTATCGGAAACCCATTTATTC-3'. The mutant sequence (underlined) is the same sequence as that used in the site-directed mutagenesis procedure performed on Sox-2<sup>-528/+238</sup>. The sequence of the wild-type FGF-4 CCAAT dsODN probe (sense strand) is: 5'-agcttCTCCTCCCCGCGGTGATTGGCAGGCGG-3'. For gel mobility shift assays and competition analyses, 15  $\mu$ g of nuclear extract protein prepared from F9 EC cells were incubated (20 min, room temperature) in a 20  $\mu$ l volume containing 50% dialysis buffer, 10  $\mu$ g bovine serum albumin, 2  $\mu$ g of p(dIdC)p(dIdC), and approx. 20 000 cpm of labeled dsODN in the absence or presence of competitor dsODN. In gel mobility supershift assays, reaction mixtures were incubated for 1 h at 4°C with the indicated antibody prior to addition of the probe. All aspects of gel mobility shift assays were repeated and similar results were obtained.

## 3. Results

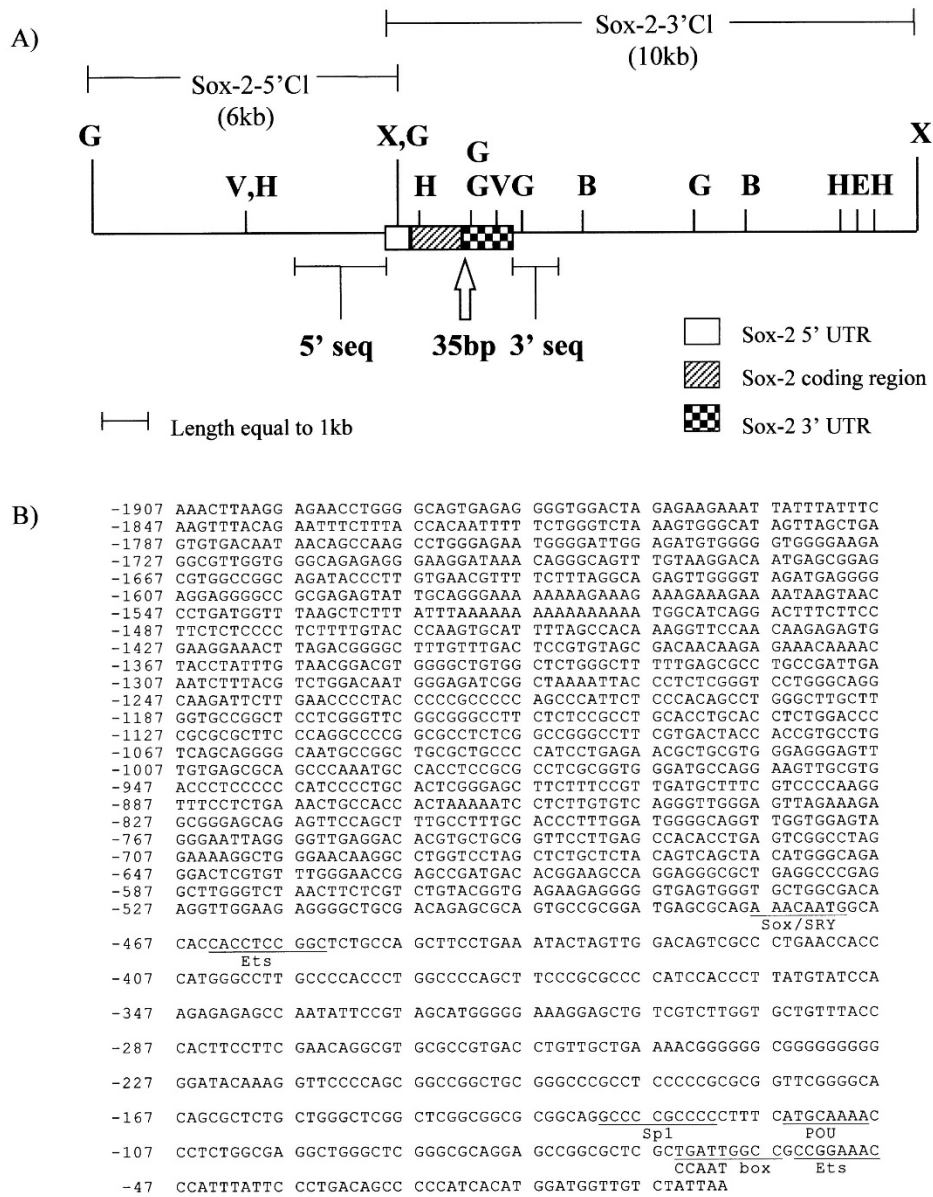
### 3.1. Isolation and structure of the murine Sox-2 gene

While homologs of the Sox-2 gene have been isolated and the cDNA sequenced in multiple species, isolation and characterization of the 5'-flanking region of Sox-2 has not been reported. Elucidation of the Sox-2 5'-flanking sequence is critical for the study of Sox-2 transcriptional regulation, as this region contains the promoter and other regulatory elements. To obtain a genomic clone containing the 5'-flanking region of the Sox-2 gene, PCR primers

were designed that amplify the region of the 5'-UTR from +17 to +183 based on the published cDNA sequence (Collignon et al., 1996). These primers were used to screen a bacteriophage P1-based genomic library from mouse strain 129 as described in section 2.1. The P1 clone was digested with various restriction enzymes and the resulting fragments cloned into pBluescript II KS+. PCR screening of the subclones indicated the presence of the *Sox-2* 5'- and 3'-flanking regions within a *Bgl*III and a *Xho*I fragment, respectively (fig. 1A), which were then designated Sox-2-5'Cl and Sox-2-3'Cl. Using deletional subcloning and primer walking, the 5'-flanking region of the *Sox-2* gene was sequenced on both strands from -1907 to -1 bp (fig. 1B, GenBank Acc. AF118260). Using similar methods, 690 bp of the 3'-flanking region of *Sox-2* was also sequenced (data not shown, GenBank Acc. AF118262) and found to exhibit greater than 80% identity to the 3'-flanking region of the sheep *Sox-2* gene (Payen et al., 1997).

While Collignon et al. (1996) reported that the coding region of the *Sox-2* gene does not contain any introns, it was unclear as to whether the untranslated regions were similarly free of introns. Using sequencing primers designed from the published cDNA sequence, Sox-2-5'Cl and Sox-2-3'Cl were screened for the presence of introns. The genomic sequence corresponding to the 5'-UTR of *Sox-2* was identical to the 5'-UTR sequence previously published, indicating that introns are absent from the 5'-UTR of *Sox-2*. However, a single 35 bp sequence was identified in the 3'-UTR that is not present in the cDNA sequence (Yuan et al., 1995; Collignon et al., 1996; Johnson et al., 1998). The 35 bp segment (data not shown, GenBank Acc. AF118261) was found to be 80% homologous to the corresponding region of the sheep sequence (Payen et al., 1997). This region could be an intron, although its ends do not conform to the classical splice acceptor or donor consensus sequences. It is conceivable that this 35 bp sequence is the result of a sequencing error of the *Sox-2* cDNA, due to the long run of adenine residues flanking both ends of the 35 bp sequence. However, it is unlikely that the 35 bp segment is an allelic variation. The sequence reported by Yuan et al. (1995) was obtained from *Sox-2* cDNA derived from F9 EC cells. This cell line and the genomic clone sequenced in the present study were both derived from mouse strain 129.

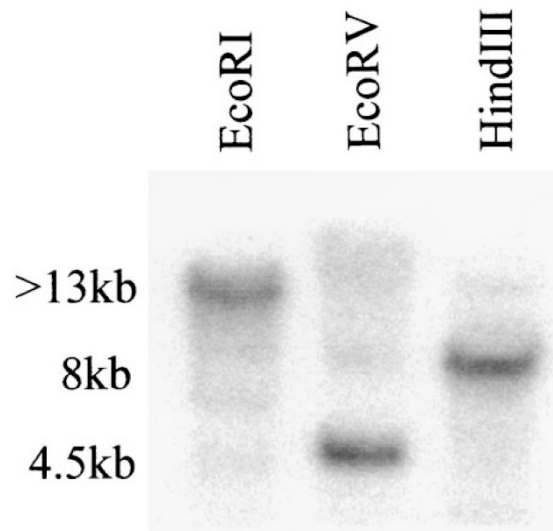




**Figure 1.** (A) Restriction enzyme mapping of the murine *Sox-2* gene. Restriction enzyme abbreviations are: G = *Bgl*II, V = *Eco*RV, H = *Hind*III, X = *Xho*I, B = *Bam*HI, E = *Eco*RI. Location of 5'- and 3'-flanking regions sequenced in this report are indicated as 5' seq and 3' seq, respectively. Location of *Sox-2* 5'-UTR (untranslated region), coding region, and 3'-UTR are noted. Location of the unpublished 35 bp region in 3'-UTR is also noted. Genomic clones Sox-2-5'Cl and Sox-2-3'Cl were isolated by restriction digest of the *Sox-2* genomic fragment with *Bgl*II and *Xho*I, respectively, and cloned into pBluescript II KS+. These two genomic clones overlap each other by 40 bp. (B) Sequence of murine *Sox-2* 5'-flanking regions from -1907 to -1 (GenBank Acc. No. AF118260). The sequence numbering is based upon defining the 5' bp of the cDNA sequence reported by Collignon et al. (1996) as the

+1 coordinate. Thus, the first base in the sequence reported by Collignon et al. (+1) immediately follows base -1 in our sequence. Putative transcription factor binding sites (underlined) include: an inverted CCAAT box/NF- $\gamma$  motif centered at -60, two Ets sites centered at -50 and -460, and an Sp1 site centered at -125. Additionally, there is a putative POU site centered at -110 and Sox/SRY site at -475.

We also performed Southern blot analysis of the murine *Sox-2* gene using genomic DNA from murine ES cells. The DNA was digested with either *EcoRI*, *EcoRV*, or *HindIII*, transferred onto a nylon membrane, and probed with a 1 kb fragment from a *Sox-2* cDNA clone. In all cases a single band was detected (fig. 2). The fragment bands were 4.5 kb following *EcoRV* digestion and 8 kb following *HindIII* digestion. These results correspond to the sizes predicted from sequence data and restriction enzyme mapping of Sox-2-5'Cl and Sox-2-3'Cl (fig. 1A). A band >13 kb was detected following *EcoRI* digestion, which also was detected with a probe located 5 kb downstream from the *Sox-2* transcript (data not shown). These findings argue that the murine genome does not contain a *Sox-2* pseudogene.

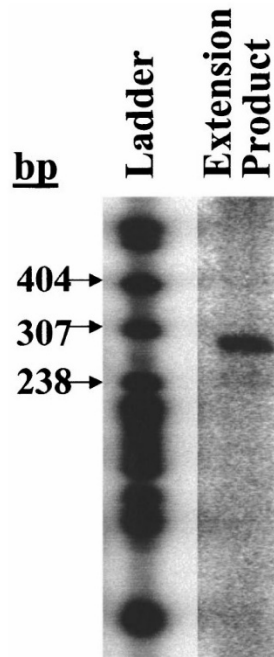


**Figure 2.** Southern blot analysis of the murine *Sox-2* gene. The Southern blot analysis was performed using 10  $\mu$ g of genomic DNA isolated from R1 ES cells digested with *EcoRI*, *EcoRV*, or *HindIII* and transferred onto a nylon membrane. The blot was probed with a radiolabeled *HindIII* fragment from a *Sox-2* cDNA clone as described in section 2.3.

### 3.2. Mapping the transcriptional start site of the *Sox-2* gene

The cDNA sequence of the *Sox-2* gene reported by Collignon et al. (1996) contains a 406 bp 5'-UTR. However, it has yet to be demonstrated as to whether the 5'-end of this sequence represents the primary start site for *Sox-2* transcription or if the gene utilizes multiple sites of transcription initiation. To determine the location of the primary transcription start site of the *Sox-2* gene, primer extension analysis was performed as described in section 2.4 us-

ing poly(A)<sup>+</sup> RNA isolated from F9 EC cells. Using the primer Sox2PE, a single major product was detected that migrates with a length of 260–280 bases (fig. 3), which corresponds well with the length of 269 bases predicted from the cDNA sequence reported by Collignon et al. (1996). While these data do not discount the possibility that there may be other minor transcription start sites utilized by the *Sox-2* promoter, they do indicate that the *Sox-2* gene initiates transcription at a single major start site approx. 406 bp upstream of the *Sox-2* protein coding region.

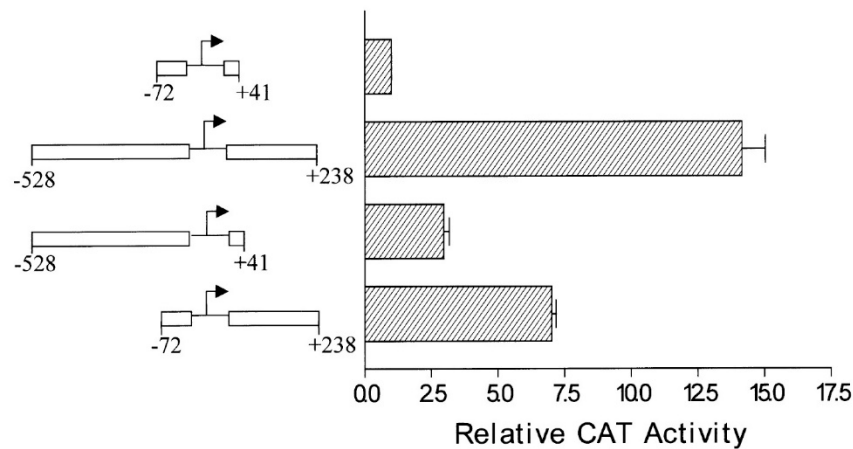


**Figure 3.** Primer extension analysis to identify the *Sox-2* transcription start site. Three micrograms of poly(A)<sup>+</sup> RNA isolated from F9 EC cells was hybridized to the radiolabeled antisense primer Sox2PE, and a primer extension reaction was performed using Moloney murine leukemia virus reverse transcriptase as described in section 2.4. The resulting cDNA product was electrophoresed on a 6% polyacrylamide–7 M urea gel adjacent to a radiolabeled pBR322-*MspI* digest and visualized using a PhosphorImager (Molecular Dynamics, Sunnyvale, California). This experiment was repeated and similar results were obtained.

### 3.3. Characterization of the *Sox-2* promoter

To begin the characterization of the *Sox-2* promoter, a panel of promoter/reporter gene constructs was generated that contain overlapping segments of the 5'-flanking region of the *Sox-2* gene placed upstream of a *cat* reporter gene as described in Section 2.6 (fig. 4). To search for positive regulatory elements present within and near the *Sox-2* promoter, these constructs were transiently transfected into F9 EC cells, which express the endogenous *Sox-2* gene. The relative *cat* expression driven by each of the constructs was normalized using

a vector containing the  $\beta$ Gal gene as an internal standard to adjust for differences in transfection efficiency. The expression of Sox-2 $_{-72/+41}$ , the smallest construct, was set arbitrarily to 1 for the purpose of comparison. The region of Sox-2 between -528 and +238 clearly contains positive *cis*-regulatory elements. The expression of Sox-2 $_{-528/+238}$  is approx. 14-fold above that of the basal construct Sox-2 $_{-72/+41}$  (fig. 4). In addition, Sox-2 $_{-528/+41}$  and Sox-2 $_{-72/+238}$  both contain positive regulatory elements, as they are able to activate the *cat* reporter gene 3-fold and 7-fold, respectively, above that of Sox-2 $_{-72/+41}$ . These results argue that the Sox-2 promoter region contains at least two positive *cis*-regulatory elements, present between the regions of (-528 to -72) and (+41 to +238), which are able to elevate expression of the reporter gene independent of one another (fig. 4).

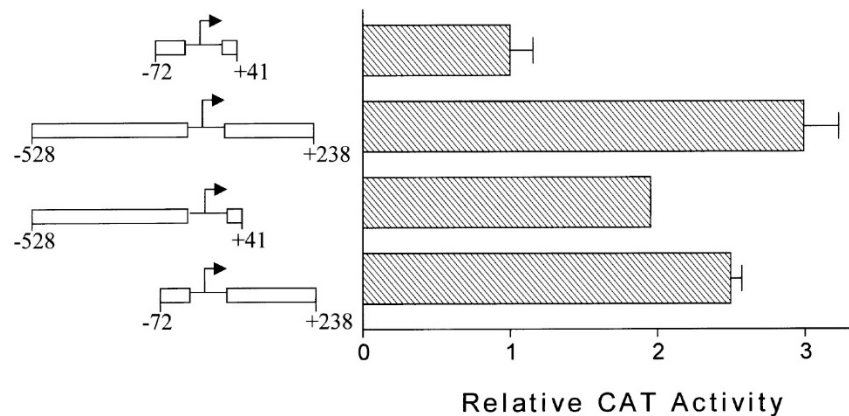


**Figure 4.** Murine Sox-2 promoter activity in F9 EC cells. Duplicate plates of F9 EC cells were transfected with the indicated Sox-2 promoter/reporter gene construct (left) and analyzed for promoter activity as a function of CAT activity (right). Any difference in the relative CAT activity between the duplicate plates assayed is shown as an error bar, which represents 1 S.D. Units of relative CAT activity of the Sox-2 $_{-72/+41}$  construct have been assigned the value of 1. In the experiment shown, CAT activity of Sox-2 $_{-72/+41}$  was 1200 cpm. The cells were also transfected with SV40- $\beta$ Gal to adjust for any differences in transfection efficiency. This experiment was performed three times and similar results were obtained in each experiment.

### 3.4. Effect of differentiation on the Sox-2 promoter constructs

In the presence of retinoic acid, F9 EC cells differentiate into cells that exhibit the properties of parietal extra-embryonic endoderm (Strickland and Mahdavi, 1978). A comparison of mRNA expression between parental and differentiated F9 EC cells revealed that expression of the Sox-2 gene becomes undetectable following differentiation (Yuan et al., 1995; Johnson et al., 1998). To determine whether the Sox-2 promoter/reporter gene constructs are differentially regulated, the Sox-2 promoter/reporter constructs were transfected into F9 EC-differentiated cells. In the EC-differentiated cells, the expression of the *cat* reporter when driven by Sox-2 $_{-528/+238}$  is approx. 3-fold greater than the basal Sox-2 $_{-72/+41}$  construct (fig. 5). This is in contrast to the activity of the same promoter region in the parental cells,

where a 14-fold induction is observed (fig. 4). Down-regulation of the two smaller constructs was also observed when compared to their expression in the parental cells, although not as dramatic as the decrease observed with Sox-2<sup>-528/+238</sup>. These results indicate that the Sox-2 promoter/reporter constructs contain *cis*-regulatory elements that are differentially utilized in F9 EC cells and their differentiated counterparts, and for the most part mimic the expression pattern of the endogenous Sox-2 gene (Yuan et al., 1995; Johnson et al., 1998). It is worth noting that the Sox-2<sup>-528/+238</sup> promoter/reporter gene construct exhibited some reporter activity in the F9 EC-differentiated cells (fig. 5). However, Sox-2 transcripts are not detected by Northern blot analysis of RNA prepared from differentiated EC cells (Yuan et al., 1995; Johnson et al., 1998). Hence, reporter activity in the differentiated cells is unlikely to be due to residual undifferentiated EC cells. Currently, the reason for this common observation is unclear, but similar observations have been made for the FGF-4, T $\beta$ R-II, and TGF- $\beta$ 2 genes (Ma et al., 1992; Kelly et al., 1995, 1998), which are differentially expressed by F9 EC cells and their differentiated counterparts.

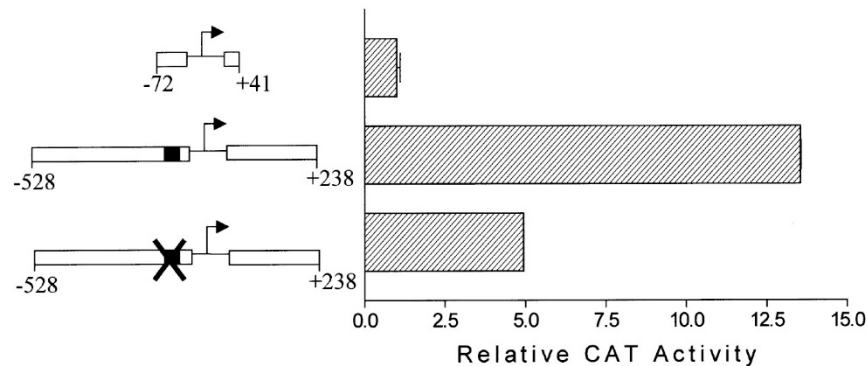


**Figure 5.** Murine Sox-2 promoter activity in F9 differentiated EC cells. Duplicate plates of F9 EC cells were induced to differentiate in 5  $\mu$ M retinoic acid. After 72 h, the differentiated cells were transfected with the indicated Sox-2 promoter/reporter gene construct (left) and analyzed for promoter activity as a function of CAT activity (right). Any difference in the relative CAT activity between the duplicate plates assayed is shown as an error bar, which represents 1 S.D. Units of relative CAT activity of the Sox-2<sup>-72/+41</sup> construct have been assigned the value of 1. In the experiment shown, CAT activity of Sox-2<sup>-72/+41</sup> was 1183 cpm. The cells were also transfected with SV40- $\beta$ Gal to adjust for any differences in transfection efficiency. This experiment was performed three times and similar results were obtained in each experiment.

### 3.5. Role of a consensus CCAAT box motif present in the Sox-2 promoter

To identify putative regulatory factors that may act on the Sox-2 promoter, we analyzed the sequence between -528 and +238 using a database containing known transcription factor binding sites (TFSEARCH). Although this analysis revealed a number of potential transcription factor binding sites (fig. 1B), no identifiable TATA box was found. However, as in the case of many promoters that lack a TATA box, an inverted CCAAT box is present.

This CCAAT box motif is located at  $-60$ , which has been found to be the average distance from the start site for functional CCAAT boxes in promoters that lack a TATA box (Mantovani, 1998). To examine whether the putative CCAAT box motif was functional, site-directed mutagenesis was used to scramble the sequence of the putative CCAAT box in Sox-2<sup>-528/+238</sup>. The ability of the mutant and its wild-type counterpart to drive expression of the *cat* reporter gene was compared in undifferentiated F9 EC cells. Mutagenesis of the CCAAT box in Sox-2<sup>-528/+238</sup> reduced expression of the reporter gene approx. 3-fold when compared to the wild-type expression (fig. 6). This result argues strongly that the CCAAT box motif present in the Sox-2 promoter plays a functional role in the transcription of this gene.



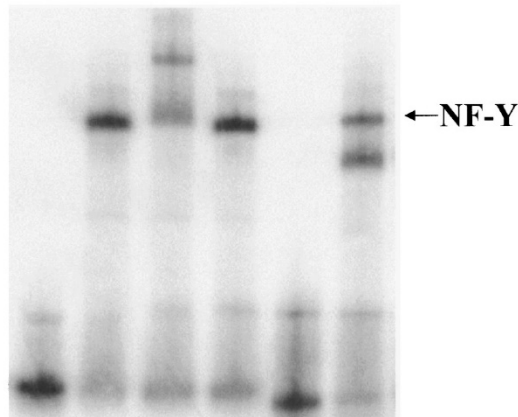
**Figure 6.** Role of the CCAAT box motif in the expression of the Sox-2 promoter in EC cells. Duplicate plates of F9 EC cells were transfected with either the wild-type or mutagenized Sox-2 promoter/reporter gene construct (left) and analyzed for promoter activity as a function of CAT activity (right). Any difference in the relative CAT activity between the duplicate plates assayed is shown as an error bar, which represents 1 S.D. Units of relative CAT activity of the Sox-2<sup>-72/+41</sup> construct have been assigned the value of 1. In the experiment shown, CAT activity of Sox-2<sup>-72/+41</sup> was 1185 cpm. The cells were also transfected with SV40- $\beta$ Gal to adjust for any differences in transfection efficiency. This experiment was performed three times and similar results were obtained in each experiment.

### 3.6. The CCAAT box present in the Sox-2 promoter binds the transcription factor NF-Y

CCAAT box motifs within promoters have been shown to bind a number of different transcription factors, including NF-Y, CDP, and CTF/NF-I (Mantovani, 1998). To help identify the transcription factor that binds this site, we analyzed the sequence flanking the CCAAT box in the Sox-2 promoter and found that it was almost identical to a 10 bp sequence found in the promoters of the FGF-4, type II transforming growth factor- $\beta$  receptor (*T $\beta$ R-II*), and cytochrome P4501A1 (*CYP1A1*) gene. These genes all contain the sequence (5'-TGATTGGCAG-3'), which differs by only one base pair (underlined) with the Sox-2 CCAAT box (5'-TGATTGGCCG-3'). Previous gel mobility shift assays demonstrated that the CCAAT box motif from each of these three genes can form two DNA/protein complexes, one of which contains NF-Y (Boucher et al., 1993, 1995; Lamb et al., 1997; Kelly et al., 1998). To determine whether NF-Y binds the Sox-2 CCAAT box motif, we performed gel mobility shift analysis

using a dsODN probe containing the CCAAT box sequence of the *Sox-2* promoter. This labeled DNA fragment was incubated with nuclear extract prepared from F9 EC cells and subjected to electrophoresis in a high ionic strength non-denaturing polyacrylamide gel. We determined that a single distinct DNA/protein complex is formed (fig. 7), which could be competed with 100-fold excess of unlabeled wild-type competitor (data not shown). However, the complex was not competed by 100-fold excess of a mutant competitor, which contained a scrambled CCAAT box sequence (data not shown). Additionally, the *Sox-2* CCAAT box motif did not form a faster migrating complex, which formed with the probe contained the *FGF-4* CCAAT box motif (fig. 7). To identify the factor in the DNA/protein complex, the labeled wild-type *Sox-2* CCAAT probe was incubated with F9 EC nuclear extract in the presence or absence of antibody specific for the A subunit of NF-Y or a control IgG antibody. Addition of the NF-YA antibody resulted in a supershift of the DNA/protein complex, whereas the control IgG antibody did not affect the complex (fig. 7). This indicates that the functional CCAAT box motif present in the *Sox-2* promoter can bind NF-Y in vitro.

	<b>Sox-2</b>				<b>FGF-4</b>	
	<b>CCAAT box</b>				<b>CCAAT box</b>	
F9 Extract	-	+	+	+	-	+
NF-YA Ab			+			
IgG Ab				+		



**Figure 7.** NF-Y binds to the CCAAT box motif present in the *Sox-2* promoter. Prior to incubation with  $^{32}\text{P}$ -labeled *Sox-2* or *FGF-4* dsODN, nuclear extracts prepared from F9 EC cells were incubated in the absence or presence of an NF-YA antibody or a nonspecific IgG antibody as indicated. Binding reactions to the labeled dsODN were performed as described in section 2.7. This experiment was repeated and similar results were obtained.

#### 4. Discussion

The present study reports the isolation and characterization of the *Sox-2* promoter. We have mapped the major transcriptional start site, using primer extension analysis, to approx. 406 bp upstream of the translation start site, which corresponds to the 5' base of the Collignon et al. (1996) *Sox-2* cDNA sequence and the -1 coordinate reported in this study (fig. 1B). Nearly 2 kb of the 5'-flanking region upstream of this transcription start site was sequenced. Utilizing this sequence, four overlapping segments of the 5'-flanking region were used to generate a panel of *Sox-2* promoter/reporter gene constructs. This resulted in the identification of two positive regulatory regions present within the *Sox-2* promoter region between -528 and +238. Furthermore, we determined that the *Sox-2* promoter/reporter gene constructs are expressed at significantly lower levels in differentiated EC cells when compared to their undifferentiated counterparts. This suggests that the regulation of the endogenous *Sox-2* gene observed during the differentiation of EC cells is likely to be due, at least in part, to differential utilization of cis-regulatory elements in close proximity to the *Sox-2* promoter.

To further define the *cis*-regulatory elements of this gene, the TFSEARCH database was used to examine the 5'-flanking sequence of the *Sox-2* gene for possible transcription factor binding sites. This analysis identified a putative CCAAT box motif centered at -60. Mutagenesis of this CCAAT box motif demonstrated that it plays a functional role in the regulation of the *Sox-2* promoter and gel mobility shift analysis demonstrated that the CCAAT box motif can bind the transcription factor NF-Y in vitro. NF-Y (also known as CBF) is a ubiquitously expressed, trimeric transcription factor that binds to CCAAT motifs in a number of eukaryotic promoters that are expressed in a tissue-specific manner (Mantovani, 1998). NF-Y has also been proposed to play a pivotal role in bridging upstream factors with the basal transcription machinery in promoters similar to that of *Sox-2*, which contain no identifiable TATA box. Interestingly, the *FGF-4*, *TβR-II*, and *CYP1A1* genes each contain identical CCAAT box-like sequence motifs in each of their respective promoters (5'-TGATTGGCAG-3'). Unlike the *Sox-2* CCAAT box motif, gel mobility shift analysis of the *FGF-4*, *TβR-II*, and *CYP1A1* CCAAT box motifs revealed formation of two distinct DNA/protein complexes, one of which contains NF-Y (Boucher et al., 1993, 1995; Lamb et al., 1997; Kelly et al., 1998). Both complexes are specific for the CCAAT box-like motif, as an unlabeled *FGF-4*, *TβR-II*, or *CYP1A1* mutant CCAAT box competitor, in which only the CCAAT box has been scrambled, does not compete for either complex (Kelly and Rizzino, unpublished data). In this regard, it should be noted that the flanking sequences of these probes appear to play no significant role in the formation of the DNA/protein complexes, as the flanking sequence in each of these probes varies considerably. Interestingly, the transcription factor(s) in the faster migrating complex, which has yet to be identified, does not bind to the *Sox-2* CCAAT box in vitro (fig. 7). This observation reveals that the unknown factor exhibits high sequence specificity, as it binds to the *FGF-4*, *TβR-II*, and *CYP1A1* CCAAT box motif (5'-TGATTGGCAG-3'), but not the *Sox-2* promoter CCAAT box motif (5'-TGATTGGCCG-3') which differ by only 1 bp (underlined) within the CCAAT box motif. Studies performed by Piechocki et al. (1998) have also demonstrated that alteration of



this base pair within the *CYP1A1* CCAAT box decreases binding of a faster migrating complex, lending further support to the critical affect of the 1 bp difference noted above.

In addition to the CCAAT box motif, a number of potential regulatory elements were identified by the use of the TFSEARCH database, including: Sp1, Sox/SRY, POU, and Ets family members. A sequence centered at -50 (GCCGGAAACC) next to the CCAAT box/NF-Y site is similar to binding sites reported for various Ets family members, as is a sequence centered at -460 (inverted, GCCGGAGGTG) (Wasylyk et al., 1993). The latter putative Ets site is close to a Sox/SRY/HMG motif centered at -475, which has a sequence (AACAAATG) that may bind SRY or several other members of the Sox family (Wegner, 1999). A sequence centered at -125 has a strong fit to the Sp1 consensus sequence (Bucher, 1990). The neighboring sequence centered at -110 (ATGCAAAA) is similar to the POU site sequence ATGCAAAT, which can bind octamer binding proteins (Okamoto et al., 1990). Within the 5'-UTR of the *Sox-2* gene, an additional putative Sox/SRY/HMG site centered at +174 (AACAAAT) may also bind members of the Sox family (Wegner, 1999). Furthermore, a nonconsensus E2F site (TTTGGAGC) centered at +134 and nonconsensus SRY site (AACTAA) centered at +163 were also identified in the 5'-UTR by the TFSearch database. Further studies will be needed to determine whether any of these potential transcription factor binding motifs influence the expression of the *Sox-2* gene.

Many early developmental genes are differentially regulated at the level of transcription. The expression pattern of *Sox-2* in EC cells also appears to follow this trend, since *Sox-2* mRNA levels become undetectable following EC cell differentiation (Yuan et al., 1995; Johnson et al., 1998). The expression pattern exhibited by the *Sox-2* promoter constructs in the EC-differentiated cells indicates that this portion of the *Sox-2* gene contains differentially utilized *cis*-regulatory elements. The putative HMG, POU, and Ets bindings sites present in the *Sox-2* 5'-flanking region may bind transcription factors that are known to be differentially regulated in EC cells (Xin et al., 1992; Yuan et al., 1995). Hence, one or more of these potential candidate transcription factor binding sites could contribute to the differential expression of *Sox-2* in EC cells and EC-derived differentiated cells. Additionally, it should be noted that none of the *Sox-2* promoter/reporter gene constructs was completely silent in the differentiated EC cells. This hints that perhaps other distal regulatory elements, which are not present in our promoter/reporter gene constructs, may also silence the endogenous *Sox-2* gene upon differentiation of EC cells.

Methylation of DNA rich in guanine and cytosine content represents another mechanism by which gene expression may be influenced (Razin, 1998). Methylated CpG islands are capable of binding multicomponent repressor complexes, which are capable of maintaining histones in a deacetylated state, thus silencing a gene by retaining the local chromatin structure in a tightly closed conformation (Razin, 1998). The region of the *Sox-2* gene from -1907 to -1200 bp upstream of the transcriptional start site and the 3'-UTR contain only 3% CpG content. In contrast, a CpG island covers the first 1200 bp upstream of the transcription start site (7% CpG) as well as *Sox-2* coding sequence and 5'-UTR (9% CpG each). The greatest CpG content is centered over the translational initiation codon (15% CpG in 300 bp) and immediately upstream of the transcriptional start site (11% CpG in 300 bp).

In conclusion, we describe the sequencing and initial characterization of the promoter of the murine *Sox-2* gene. Given the critical role of *Sox-2* during early development (Yuan et al., 1995; Botquin et al., 1998; Johnson et al., 1998; Nishimoto et al., 1999), the regulation of this gene warrants further analysis. This study provides a foundation for future efforts to elucidate the mechanisms by which the *Sox-2* gene is regulated both spatially and temporally.

**Acknowledgments** – Tamara Nowling is thanked for reading the manuscript and providing helpful comments. This work was supported by a grant from the National Institutes of Health (CA 74771) and by a core grant to the University of Nebraska Medical Center Eppley Cancer Center from the National Cancer Institute (Laboratory Cancer Research Center Support Grant, CA 36727).

## References

- Botquin, V., Hess, H., Fuhrmann, G., Anastassiadis, C., Gross, M.K., Vriend, G., Scholer, H.R., 1998. New POU dimer configuration mediates antagonistic control of an osteopontin preimplantation enhancer by Oct-4 and Sox-2. *Genes Dev.* 12, 2073–2090.
- Boucher, P.D., Piechocki, M.P., Hines, R.N., 1995. Partial characterization of the human CYP1A1 negatively acting transcription factor and mutational analysis of its cognate DNA recognition sequence. *Mol. Cell Biol.* 15, 5144–5151.
- Boucher, P.D., Ruch, R.J., Hines, R.N., 1993. Specific nuclear protein binding to a negative regulatory element on the human CYP1A1 gene. *J. Biol. Chem.* 268, 17384–17391.
- Bucher, P., 1990. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.* 212, 563–578.
- Collignon, J., Sockanathan, S., Hacker, A., Cohen-Tannoudji, M., Norris, D., Rastan, S., Stevanovic, M., Goodfellow, P.N., Lovell-Badge, R., 1996. A comparison of the properties of Sox-3 with Sry and two related genes, Sox-1 and Sox-2. *Development* 122, 509–520.
- Dignam, J.D., Lebovitz, R.M., Roeder, R.G., 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* 11, 1475–1489.
- Fried, M., Crothers, D.M., 1981. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res.* 9, 6505–6525.
- Johnson, L.R., Lamb, K.A., Gao, Q., Nowling, T.K., Rizzino, A., 1998. Role of the transcription factor Sox-2 in the expression of the FGF-4 gene in embryonal carcinoma cells. *Mol. Reprod. Dev.* 50, 377–386.
- Kamachi, Y., Sockanathan, S., Liu, Q., Breitman, M., Lovell-Badge, R., Kondoh, H., 1995. Involvement of SOX proteins in lens-specific activation of crystallin genes. *EMBO J.* 14, 3510–3519.
- Kelly, D., Kim, S.J., Rizzino, A., 1998. Transcriptional activation of the type II transforming growth factor-beta receptor gene upon differentiation of embryonal carcinoma cells. *J. Biol. Chem.* 273, 21115–21124.
- Kelly, D., Scholtz, B., Orten, D.J., Hinrichs, S.H., Rizzino, A., 1995. Regulation of the transforming growth factor-beta 2 gene promoter in embryonal carcinoma cells and their differentiated cells: Differential utilization of transcription factors. *Mol. Reprod. Dev.* 40, 135–145.
- Kuhlbrodt, K., Herbarth, B., Sock, E., Enderich, J., Hermans-Borgmeyer, I., Wegner, M., 1998a. Cooperative function of POU proteins and SOX proteins in glial cells. *J. Biol. Chem.* 273, 16050–16057.

- Kuhlbrodt, K., Herbarth, B., Sock, E., Hermans-Borgmeyer, I., Wegner, M., 1998b. Sox10, a novel transcriptional modulator in glial cells. *J. Neurosci.* 18, 237–250.
- Lamb, K.A., Johnson, L.R., Rizzino, A., 1997. NF-Y binds to the CCAAT box motif of the FGF-4 gene and promotes FGF-4 expression in embryonal carcinoma cells. *Mol. Reprod. Dev.* 48, 301–309.
- Ma, Y.G., Rosfjord, E., Huebert, C., Wilder, P., Tiesman, J., Kelly, D., Rizzino, A., 1992. Transcriptional regulation of the murine k-FGF gene in embryonic cell lines. *Dev. Biol.* 154, 45–54.
- Mantovani, R., 1998. A survey of 178 NF-Y binding CCAAT boxes. *Nucleic Acids Res.* 26, 1135–1143.
- Nishimoto, M., Fukushima, A., Okuda, A., Muramatsu, M., 1999. The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. *Mol. Cell Biol.* 19, 5453–5465.
- Nowling, T., Johnson, L.R., Wiebe, M., Rizzino, A., 2000. Identification of the transactivation domain of the transcription factor Sox-2 and an associated co-activator. *J. Biol. Chem.* 275, 3810–3818.
- Okamoto, K., Okazawa, H., Okuda, A., Sakai, M., Muramatsu, M., Hamada, H., 1990. A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. *Cell* 60, 461–472.
- Payen, E., Pailhoux, E., Gianquinto, L., Hayes, H., Le Pennec, N., Bezard, J., Cotinot, C., 1997. The ovine SOX2 gene: Sequence, chromosomal localization and gonadal expression. *Gene* 189, 143–147.
- Pevny, L.H., Sockanathan, S., Placzek, M., Lovell-Badge, R., 1998. A role for SOX1 in neural determination. *Development* 125, 1967–1978.
- Piechocki, M.P., Hines, R.N., 1998. Functional characterization of the human CYP1A1 negative regulatory element: Modulation of Ah receptor mediated transcriptional activity. *Carcinogenesis* 19, 771–780.
- Razin, A., 1998. CpG methylation, chromatin structure and gene silencing—a three-way connection. *EMBO J.* 17, 4905–4908.
- Rizzino, A., 1989. Use of embryonal carcinoma cells to study growth factors during early mammalian development. In: Rosenblum, L.Y., Heyner, S. (Eds.), *Growth Factors in Mammalian Development*. CRC Press, Boca Raton, FL, pp. 113–134.
- Rosfjord, E., Lamb, K., Rizzino, A., 1994. Cryptic promoter activity within the backbone of a plasmid commonly used to prepare promoter/reporter gene constructs. *In Vitro Cell Dev. Biol. Anim.* 30A, 477–481.
- Scholtz, B., Lamb, K., Rosfjord, E., Kingsley, M., Rizzino, A., 1996. Appearance of nuclear protease activity after embryonal carcinoma cells undergo differentiation. *Dev. Biol.* 173, 420–427.
- Strickland, S., Mahdavi, V., 1978. The induction of differentiation in teratocarcinoma stem cells by retinoic acid. *Cell* 15, 393–403.
- van de Wetering, M., Oosterwegel, M., van Norren, K., Clevers, H., 1993. Sox-4, an Sry-like HMG box protein, is a transcriptional activator in lymphocytes. *EMBO J.* 12, 3847–3854.
- Wasylyk, B., Hahn, S.L., Giovane, A., 1993. The Ets family of transcription factors. *Eur. J. Biochem.* 211, 7–18.
- Wegner, M., 1999. From head to toes: The multiple facets of Sox proteins. *Nucleic Acids Res.* 27, 1409–1420.
- Xin, J.H., Cowie, A., Lachance, P., Hassell, J.A., 1992. Molecular cloning and characterization of PEA3, a new member of the Ets oncogene family that is differentially expressed in mouse embryonic cells. *Genes Dev.* 6, 481–496.

Yuan, H., Corbi, N., Basilio, C., Dailey, L., 1995. Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes Dev.* 9, 2635–2645.