

Trans-activation and DNA-binding properties of the transcription factor, Sox-18

Brett M. Hosking, George E. O. Muscat*, Peter A. Koopman, Dennis H. Dowhan and Timothy L. Dunn

Centre for Molecular and Cellular Biology, University of Queensland, St Lucia, 4072 Queensland, Australia

Received April 25, 1995; Revised and Accepted June 14, 1995

GenBank accession no. L35032

ABSTRACT

Sox-18 is a member of the Sox multi-gene family (Sry-related HMG-box gene). We have bacterially expressed this 378 amino acid protein and demonstrated sequence-specific binding to the Sox DNA-binding motif AACAAAG. A distinct 95 amino acid activation domain was mapped in Sox-18 using GAL4–Sox-18 fusions (amino acids 160–225). Furthermore, Sox-18 was capable of trans-activating gene expression through the AACAAAG motif. Our results suggest that Sox-18 functions as a classical trans-activator of gene expression.

The discovery in 1990 of the murine sex determining gene, *Sry* (1,2) rapidly led to the identification of the *Sox* gene family, encoding proteins which share homology in their DNA-binding domain (3–8). This DNA-binding domain is closely related to that of the nuclear proteins known as the high mobility group (HMG) proteins. The HMG-box DNA-binding domain is ~80 amino acids and contains highly conserved proline, aromatic and basic residues (see 9 for review). The *Sox* gene family consists of at least 18 different proteins in the mouse, with orthologues across the plant and animal kingdoms (9). All *Sox* genes analysed to date encode proteins which bind specifically to the double-stranded DNA motif, $A/T^A/TCAA^A/TG$ (4,5,9–12). A clear biological role has been assigned only to the *Sry* protein, which is necessary and sufficient for male development of chromosomally female (XX) mice (2). The other members of the *Sox* family are less well understood, although those that have been at least partially characterised, show definite tissue- and stage-specific expression. *TCF-1* (13), *LEF-1* (3) and *Sox-4* (5), have been implicated in the regulation of lymphoid differentiation and have been shown to be restricted in their expression, *Sox-5* is expressed in post-meiotic round spermatids, suggesting a role in spermatogenesis (4,10) and *Sox-9* is expressed during chondrogenesis in mouse embryos (14). We have recently cloned and sequenced a new member of this gene family, *Sox-18* which is expressed only in lung, heart and skeletal muscle tissues of the adult mouse (15).

We have expressed recombinant Sox-18 in *Escherichia coli* as a glutathione-S-transferase (GST) fusion protein. This recombinant Sox-18 interacts with the Sox Consensus Motif (SoCM = AACAAAG) in a dose-dependent manner (data not shown).

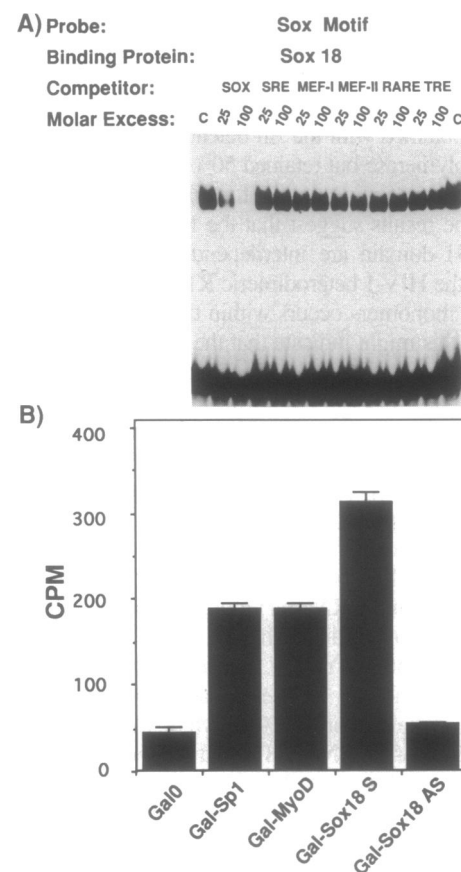


Figure 1. (A) EMSA of ^{32}P kinase radio-labelled SoCM (0.5 ng) and affinity purified recombinant GST-Sox-18 fusion protein (5 μ g) performed as previously described (19). Competition with unlabelled double-stranded oligo was carried out at 25- and 100-fold molar excess. C denotes the control binding reaction in the absence of unlabelled competitor. The unlabelled competitors are SoCM, non-hormone response elements (SRE, MEF-I and MEF-II) and the characterised steroid hormone response elements [mCRBPI (RARE) and α MHC (TRE)]. (B) Transcriptional activation of Gal4 chimeric constructs in transient transfection assays into COS-1 cells using DOTAP (Boehringer-Mannheim). Five μ g of the pG5E1bCAT reporter (containing five Gal4 binding motifs) was co-transfected with 3 μ g of each of the Gal4 chimeras indicated. Gal0 contains amino acids 1–147 of the yeast Gal4 transcription factor responsible for the DNA binding activity, Gal-Sp1 and Gal-MyoD contain the complete ORF of the Sp1 and MyoD transcription factors respectively.

* To whom correspondence should be addressed

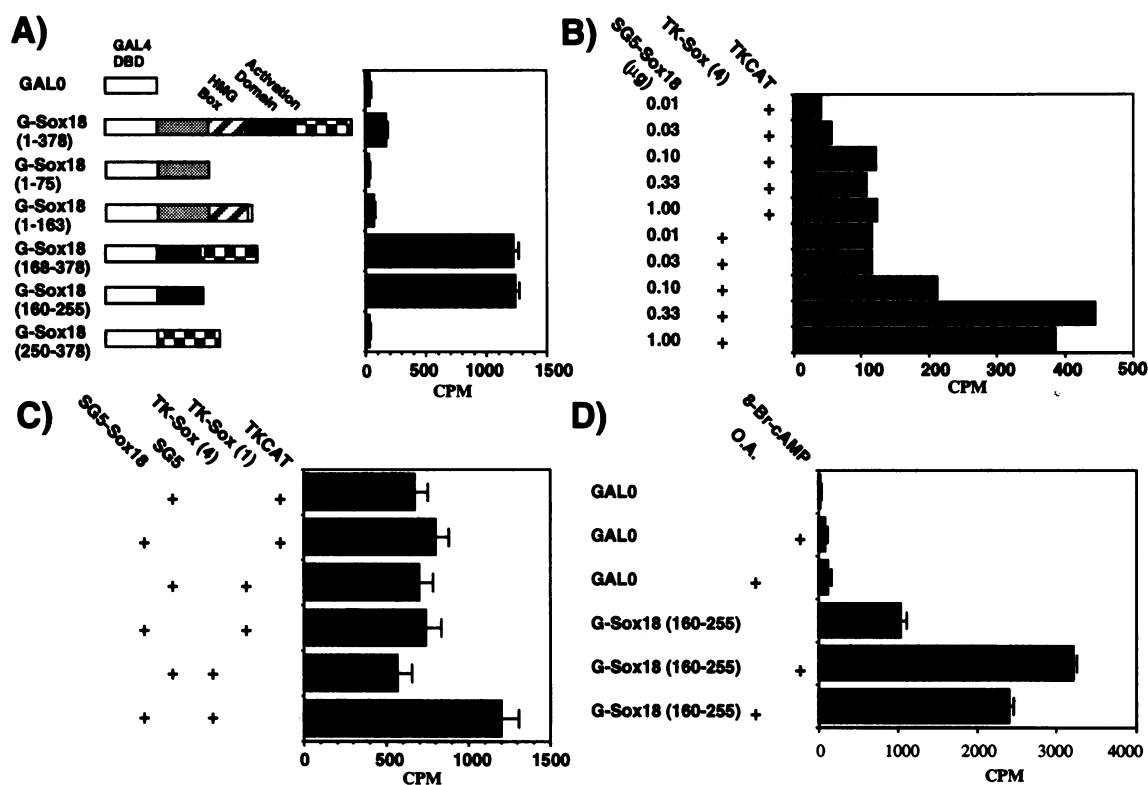


Figure 2. (A) Gal-Sox18 expression constructs used in co-transfection assays to map the activation domain of Sox-18 (performed as in Fig. 1). The respective level of activation of the pGSE1bCAT reporter is shown. The G refers to the plasmid GAL0 into which the various fragments were cloned, the numbers in brackets correspond to the amino acid positions in murine Sox-18. (B) Co-transfection assays of the eukaryotic expression plasmid pSG5-Sox18 and TK-Sox(4)CAT, containing four Sox consensus motifs cloned into TK-CAT. Five µg of CAT reporter was co-transfected with increasing amounts of SG5-Sox18. (C) Co-transfection of 5 µg of TK-CAT, TK-Sox(1) (containing one Sox motif) or TK-Sox(4) (containing four Sox motifs) with the 2 µg of control vector (SG5) or the Sox-18 expression vector (SG5-Sox-18). (D) The effect of phosphorylation reagents 8-bromo cAMP (0.5 mM) and okadaic acid (50 nM) on the transcriptional activity of the Sox-18 activation domain. CAT assays were performed on COS-1 cells co-transfected with either Gal0 or G-Sox18 (160–255) and treated with either 8-bromo cAMP or okadaic acid. Results shown are the mean CAT activity and the SDs derived from experiments done at least in triplicate.

EMSA competition assays indicated that only the unlabelled SoCM specifically competed the Sox-18–SoCM complex, none of the non-HMG-box binding sites had any effect (Fig. 1A). Several other Sox-like DNA motifs (specifically AACAAATC, GACAAAG, TACAATC, AACAAAC and CACAATTG) were also unable to effectively compete the Sox-18–SoCM complex (data not shown). We therefore conclude Sox-18 binds DNA in a highly sequence-specific manner, though without rigorous binding site selection experiments there remains the formal possibility that Sox-18 also binds another unrelated sequence.

Analysis of transcription factors using the Gal4 hybrid assay system, where chimeric genes are constructed using the Gal4 DNA-binding domain and protein domains of choice, has been widely utilised to determine whether the chosen protein is capable of regulating transcription and for the identification of modular domains (5,11,16,17). We initially constructed a Gal-Sox18 fusion containing the full ORF of *Sox-18*. Co-transfection of this Gal-Sox18 fusion with pGSE1bCAT, which contains five Gal4 target sites (18), into COS-1 cells demonstrated that Sox-18 is able to activate transcription 6-fold above the negative controls Gal0 and Gal-Sox18 antisense and similar to the muscle-specific MyoD and the ubiquitous Sp1 transcription factors (Fig. 1B). Using this Gal4 hybrid assay system, the activation domain of Sox-18 was mapped to a 95 amino acid segment immediately

downstream of the HMG box (amino acids 160–255, Fig. 2A). The 95 amino acid domain increased CAT activity >100-fold above background while the full length Sox-18 protein activates only ~10-fold, suggesting that the additional DNA-binding domain suppresses *trans*-activation. The activation domain of the related Sry protein was shown to be confined to a glutamine/histidine rich region, just downstream of the HMG box (11). A serine rich region in the C-terminus of Sox-4 was shown to have *trans*-activation activity (5). The activation domain of Sox-18 does not show significant homology, at either the nucleotide or amino acid level, to any previously characterised protein in the sequence databases.

To determine whether the *Sox-18* gene product could regulate transcription through the SoCM, we cloned the *Sox-18* cDNA into the SV40 expression vector, pSG5 (Stratagene). This was co-transfected with TK-CAT (basal reporter) or a TK-Sox(4) reporter (TK-CAT containing four SoCMs) into COS-1 cells. A dose-dependent specific increase in CAT expression from TK-Sox(4) was demonstrated with increasing amounts of the SG5-Sox18 (Fig. 2B). We could not detect significant *trans*-activation via a single SoCM site in TK-CAT (Fig. 2C). These results show Sox-18 can specifically *trans*-activate gene expression via the AACAAAG motif.

To examine the possibility that phosphorylation plays a role in the transcriptional regulation of Sox-18, we used the reagents 8-Br-cAMP (stimulates cAMP-dependent protein kinase A pathways) and okadaic acid (inhibits serine/threonine protein phosphatases) to globally increase the level of protein phosphorylation. We co-transfected the basal vector, Gal0 and the Gal-Sox18 activation domain construct, G-Sox18 (160–255), with the pG5E1bCAT reporter, in the presence and absence of either 8-Br-cAMP or okadaic acid. Both reagents increased the expression of CAT from the basal vector Gal0, by ~2- and 3-fold respectively (Fig. 2D). However, in the presence of the chimeric G-Sox18 (160–255) plasmid, 8-Br-cAMP increased CAT expression ~3.5-fold, whereas okadaic acid increased expression only 2-fold. These results, with respect to the effect of these phosphorylating agents on basal transcription, indicate that 8-Br-cAMP increases the activity of Sox-18, suggesting that c-AMP dependent protein kinase cascades target the Sox-18 activation domain and regulate the activity of this protein.

We report here that Sox-18 binds specifically to the Sox consensus motif AACAAAG and is able to activate transcription via this site. We have mapped a novel *trans*-activation domain in the Sox-18 protein to amino acids 160–255. We are currently studying gain and loss of function of Sox-18 in various cultured cells and performing mutagenesis of the novel activation domain.

REFERENCES

- Berta, P., Hawkins, J., Sinclair, A., Taylor, A., Griffiths, B., Goodfellow, P. and Fellous, M. (1990) *Nature* **348**, 448–450.
- Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P. and Lovell-Badge, R. (1991) *Nature* **351**, 117–121.
- Travis, A., Amsterdam, A., Belanger, C. and Grosschedl, R. (1991) *Genes Dev.* **5**, 880–894.
- Denny, P., Swift, S., Connor, F. and Ashworth, A. (1992) *EMBO J.* **11**, 3705–3712.
- van de Wetering, M., Oosterwegel, M., van Norren, K. and Clevers, H. (1993) *EMBO J.* **12**, 3847–3854.
- Goze, C., Poulat, F. and Berta, P. (1993) *Nucleic Acids Res.* **21**, 2943.
- Wright, E.M., Snopek, B. and Koopman, P. (1993) *Nucleic Acids Res.* **21**, 744.
- van de Wetering, M. and Clevers, H. (1993) *Nucleic Acids Res.* **21**, 1669.
- Laudet, V., Stehelin, D. and Clevers, H. (1993) *Nucleic Acids Res.* **21**, 2493–2501.
- Connor, F., Cary, P., Read, C., Preston, N., Driscoll, P., Denny, P., Crane-Robinson, C. and Ashworth, A. (1994) *Nucleic Acids Res.* **22**, 3339–3346.
- Dubin, R.A. and Ostrer, H. (1994) *Mol. Endocrinol.* **8**, 1182–1192.
- Read, C., Cary, P., Preston, N., Lnenicek-Allen, M. and Crane-Robinson, C. (1994) *EMBO J.* **13**, 5639–5646.
- van de Wetering, M., Oosterwegel, M., Holstege, F., Dooyes, D., Suijkerbuijk, R., Geurts van Kessel A. and Clevers, H. (1992) *J. Biol. Chem.* **267**, 8530–8536.
- Wright, E., Snopek, B. and Koopman, P. (1995) *Nature Genet.* **9**, 15–20.
- Dunn, T., Mynett-Johnson, L., Wright, E., Hosking, B., Koopman, P. and Muscat, G. (1995) *Gene*, **161**, in press.
- Lee, J., Galvin, K. and Shi, Y. (1993) *Proc. Natl. Acad. Sci USA* **90**, 6145–6149.
- Weintraub, H., Dwarki, V., Verma, I., Davis, R., Hollenberg, S., Snider, L., Lassar, A. and Tapscott, S. (1991) *Genes Dev.* **5**, 1377–1386.
- Lillie, J.W. and Green, M.R. (1989) *Nature* **338**, 39–44.
- Downes, M., Griggs, R., Atkins, A., Olson, E. and Muscat, G. (1993) *Cell Growth Diff.* **4**, 901–909.