

# A conserved family of genes related to the testis determining gene, *SRY*

Paul Denny, Sally Swift, Nigel Brand<sup>1</sup>, Nina Dabhade<sup>1</sup>, Paul Barton<sup>1</sup> and Alan Ashworth\*

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

<sup>1</sup>Molecular Biology Group, Department of Cardiothoracic Surgery, NHLI, Dovehouse Street, London SW3 6LY, UK

Submitted April 28, 1992

EMBL accession nos X65653 – X65656 (incl.) and X65658 – X65668 (incl.)

*SRY* is the Y chromosomal gene which determines testis formation during mammalian embryogenesis (1). This gene encodes a potential transcription factor with a DNA binding motif known as a HMG box (2). Four murine autosomal *SRY*-related genes have been described having high homology to the HMG box region of *SRY* (3). We have cloned additional members of this gene family from phylogenetically diverse organisms.

cDNAs prepared from mouse or human RNAs or phage lysates from cDNA libraries (*Xenopus oocyte* (4) and *Drosophila* embryo (0–12 hr) (Stratagene)), were used as templates in the polymerase chain reaction (PCR), using degenerate oligonucleotide primers which corresponded to regions conserved in the known *SRY* and *SRY*-related amino-acid sequences (Figure 1). Sequencing of the cloned PCR products demonstrated that they were heterogeneous, with the clones falling into a few, distinct classes, mostly differing from those *SRY*-related clones previously described (3). All of the encoded proteins are more closely related to *SRY* than to other HMG-box proteins, including the T-cell factor TCF-1 (5) and the product of the *S. pombe* mating type gene, *Mc* (6). The *SRY*-like genes have been named *SOX* ('*SRY*-box') genes (R. Lovell-Badge, personal communication). Some of our mammalian *SOX* cDNAs encode proteins which are very similar (>90% identity) to one or more of the *Xenopus* clones. It is difficult, therefore, to determine which may be orthologues. One of the *Xenopus* cDNAs, however, differs from the mouse and human *SOX-5* sequence by only one amino acid and hence we have named it *XSox-5*. We may have also cloned the human orthologue of mouse *Sox-4* (3), as one of our cDNAs encodes a protein which is 96% identical with the mouse sequence. *SOX-5* and *-6* were both cloned from mouse and human cDNA.

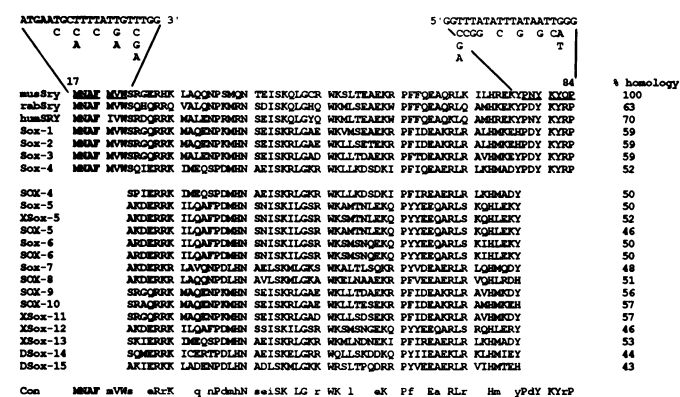
Expression of *SRY*-like genes during embryogenesis in mouse (3) and *Drosophila* and the strong evolutionary conservation of this gene family suggests that the *SRY*-related genes may be of importance in developmental processes. The *SOX* genes are highly related in their DNA binding domain, which reflects an overlap in their sequence specificity (Ref. 7 and P.D., S.S., Frances Connor and A.A., submitted). It will be important to determine the structure of these genes outside the HMG-box, as these regions may specify other functions such as protein–protein interactions.

## ACKNOWLEDGEMENTS

This research was supported by the Cancer Research Campaign, Medical Research Council and British Heart Foundation. We thank Magdi Yacoub for his support.

## REFERENCES

1. Koopman *et al.* (1991) *Nature* **351**, 117–121.
2. Jantzen *et al.* (1990) *Nature* **344**, 830–836.
3. Gubbay *et al.* (1990) *Nature* **346**, 245–250.
4. Shuttleworth *et al.* (1990) *EMBO J.* **10**, 3233–3240.
5. Van de Watering *et al.* (1991) *EMBO J.* **10**, 123–132.
6. Kelly *et al.* (1988) *EMBO J.* **7**, 1537–1547.
7. Nasrin *et al.* (1991) *Nature* **354**, 317–320.



**Figure 1.** Comparison of amino acid sequences encoded by *SRY*-related cDNAs. Sequence identity over the region encoded between the primers is shown at the right and numbering is as in Ref. 3. In the consensus sequence, residues conserved in greater than 90% of the *SOX* cDNAs are in upper case and those conserved in greater than 50% are in lower case. Human gene names are in upper case, murine in lower case; *Drosophila* genes are named *DSox*- and genes from *Xenopus*, *XSox*-. The amino acid sequences shown are each derived from several independent PCR isolates, with the exceptions of *Sox-7*, *XSox-5* and *DSox-14*. PCR reaction conditions, cloning and sequencing were as described (P.D., S.S., Frances Connor and A.A., submitted).

\* To whom correspondence should be addressed