

Seven new members of the *Sox* gene family expressed during mouse development

Edwina M.Wright, Bernadette Snopek and Peter Koopman*
Centre for Molecular Biology and Biotechnology, The University of Queensland, Brisbane, Queensland 4072, Australia

Received December 15, 1992; Accepted January 5, 1993 EMBL accession nos Z18957–Z18963 (incl.)

The mammalian Y-linked testis-determining gene *Sry* encodes a protein that contains a 79 amino acid motif, called the HMG box. This motif is responsible for sequence-specific DNA binding in *Sry* and several known transcription factors (1, 2). Several mouse genes with HMG box regions very similar to that of *Sry* have recently been described (3, 4); these have been named *Sox* 1–7 (*Sry*-type HMG box genes). Members of this gene family have also been identified in a number of other species, both vertebrate and invertebrate (4, 5, 6). Conservation of the HMG box motif across species, coupled with recent evidence that *Sox* 1–3 are involved in development of the nervous system (J.Collignon and R.Lovell-Badge, manuscript submitted) and *Sox* 5 in spermatogenesis (7), suggest important roles for this gene family in development.

We have identified an additional seven *Sox* genes in the mouse, all of which are closely related to the known *Sox* genes (Figure 1). Total RNA was prepared from 11.5 days *post coitum* (dpc) mouse embryos and reverse transcriptase polymerase chain reaction (RT-PCR) was performed using degenerate oligonucleotide primers which anneal within the HMG box (Figure 1). The PCR products were cloned and sequenced to reveal the novel genes which we have called *Sox* 8, 9, 10, 11, 12, 13 and 14. In addition to these genes, *Sox* 1–5 and *Sox* 7 were amplified in our experiments indicating that all of the known *Sox* genes (with the possible exception of *Sox* 6) are expressed in the mouse embryo at 11.5 dpc. The amino acid sequence of the *Sox* 11 fragment is identical to that of the avian gene Lf6 (6), suggesting that these two genes are orthologues. Sequence comparison of the mouse *Sox* gene family in the HMG box indicates they fall into six distinct subgroups (Figure 2); A: *Sry*; B: *Sox* 1, 2, 3 and 14; C: *Sox* 4, 11 and 12; D: *Sox* 5, 6 and 13; E: *Sox* 8, 9 and 10; F: *Sox* 7. It will be interesting to determine whether this structural classification is reflected in functional similarities among the members of the *Sox* gene family.

```
      1                                                                  79
Sry   GHVKRPMNAF MVWSRGERHK LAQQNPSMQN TEISKQLGCR WKSLTEAEKR PFFQEAQRLK TLHREKYPNY KYQPHRRAK
Sox 1 DRVKRPMNAF MVWSRGQRRK MAQENPKMHN SEISKRLGAE WKVMSEAEKR PFIDEAKRLR ALHMKEHPDY KYRPRRKTK
Sox 2 DRVKRPMNAF MVWSRGQRRK MAQENPKMHN SEISKRLGAE WKLLSETEKR PFIDEAKRLR ALHMKEHPDY KYRPRRKTK
Sox 3 DRVKRPMNAF MVWSRGQRRK MALENPKMHN SEISKRLGAD WKLLTDAEKR PFIDEAKRLR AVHMKEYPDY KYRPRRKTK
Sox 4 GHIKRPMNAF MVWSQIERRK IMEQSPDMHN AEISKRLGKR WKLLKDSDKI PFIQEAERLR LKHMADYPDY KYRPRKKVK
Sox 5 PHIKRPMNAF MVWAKDERRK ILQAFPDMHN SNISKILGSR WKAMTNLEKQ PYYEEQARLS KQHLEKYPDY KYKPRPKRT
Sox 6             ARDERRK ILQAFPDMHN SNISKILGSR WKSMSNQEKQ PYYEEQARLS KIHLEKY
Sox 7             AKDERKR LAVQNPDLHN AELSKMLGKS WKALTLSQKR PYVDEAERLR LQHMQDY

Sox 8          MVWAQAARRK LADQYPHLHN AELSKTLGKL WRLLSESEKR PFVEEAERLR VQHKKD
Sox 9          MVWAQAARRK LADQYPHLHN AELSKTLGKL WRLLNESEKR PFVEEAERLR VQHKKD
Sox10          MVWAQAAPRK LADQYPHLHN AELSKTLGKL WRLLNESDKR PFIEEAERLR MQHKKD
Sox11          MVWSKIERRK IMEQSPDMHN AEISKRLGKR WKMLKDSEKI PFIREAERLR LKHMAD
Sox12          MVWSQHERRK IMDQWPDMHN AEISKRLGRR WQLLQDSEKI PFEREAERLR LKHMAD
Sox13          MVWAKDERRK ILQAFPDMHN SSISKILGSR WKSMTNQEKQ PYYEEQARLS RQHLEK
Sox14          MVWSRGQRRK MAQENPKMHN SEISKRLGAE WKLLSEAEKR PYIDEAKRLR AQHMKE

Cons  d.vKRPMNAF MVWsrgerrk iaqqnPdmhN .eiSKrLGkr Wkll.eseKr PfieEaeRLr .qHmkdyPdY KYrPrrktk
       --------                                                          ---------
```

Figure 1. Comparison of deduced amino acid sequences of the HMG box regions of all known mouse *Sox* genes. The positions of the degenerate oligonucleotide primers used for amplification are underscored on the consensus sequence. The nucleic acid sequences of the new *Sox* genes were confirmed by analysis of amplified DNA from several independent PCRs.

Sox	2	3	4	5	6	7	8	9	10	11	12	13	14	Sry
1	**95**	**88**	55	45	49	45	54	52	50	57	54	48	**93**	63
2	.	**89**	59	43	47	47	57	55	54	59	57	46	**95**	63
3	.	.	61	43	43	49	54	54	52	61	59	46	**89**	61
4	.	.	.	45	45	57	59	59	61	**93**	**84**	48	57	54
5	.	.	.	.	**91**	47	40	40	36	49	45	**92**	47	50
6	.	.	.	.	.	40	40	38	34	47	45	**91**	49	49
7	.	.	.	.	.	.	62	62	59	59	53	45	51	47
8	.	.	.	.	.	.	.	**98**	**91**	57	61	43	57	50
9	.	.	.	.	.	.	.	.	**93**	57	61	43	55	50
10	.	.	.	.	.	.	.	.	.	55	57	39	54	46
11	.	.	.	.	.	.	.	.	.	.	**84**	52	57	54
12	.	.	.	.	.	.	.	.	.	.	.	48	55	52
13	.	.	.	.	.	.	.	.	.	.	.	.	50	52
14	.	.	.	.	.	.	.	.	.	.	.	.	.	61

Figure 2. Percentage of amino acid residues identical between members of the mouse *Sox* gene family. The comparison involved residues 11–66 as shown in Figure 1. Identities higher than 80% are in bold type, providing a basis for classification of the *Sox* genes into subgroups.

ACKNOWLEDGEMENTS

We are grateful to Jérôme Collignon and Robin Lovell-Badge for access to unpublished data.

REFERENCES

1. Harley,V. *et al.* (1992) *Science* **255**, 453–456.
2. Giese,K. *et al.* (1992) *Cell* **69**, 185–195.
3. Gubbay,J. *et al.* (1990) *Nature* **346**, 245–250.
4. Denny,P. *et al.* (1992) *Nucleic Acids Res.* **20**, 2887.
5. Picardo,A.M. *et al.* (1993) *PCR Methods. Appl.*, in press.
6. Griffiths,R. (1991) *Proc. R. Soc. Lond. B.* **244**, 123–128.
7. Denny,P. *et al.* (1992) *EMBO J.* **11**, 3705–3712.

* To whom correspondence should be addressed