1500–1501 Nucleic Acids Research, 1994, Vol. 22, No. 8

Definition of a consensus DNA binding site for SRY

Vincent R.Harley*, Robin Lovell-Badge[1] and Peter N.Goodfellow
Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH and [1]MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

Received February 17, 1994; Accepted March 15, 1994

The HMG box family is functionally diverse, with members involved in DNA recognition during nucleosome assembly, recombination and transcription (1 for review). A subclass of these proteins having a single HMG box are likely to be transcription factors and include the mammalian testis determining factor (SRY), T-cell factor (TCF-1), lymphoid enhancer factor (LEF-1 or TCF-1α) and the yeast sterility factor (Ste11); these proteins recognise not only non-B-DNA structures but also specific DNA sequences whose consensus is (A/T)(A/T)CAAA-G (2 for review). Recombinant SRY can bind these DNA sites in a sequence-dependent manner, and this activity is required for testis development (3,6). DNA binding sites for SRY have been deduced from the target sequences of related genes e.g. TCF-1 (3,4), LEF-1 (5), IRE-ABP (6) or from promoter regions of candidate downstream genes e.g. *AMH* and *aromatase* (7,8), but it remains to be determined whether these are optimal sites for interaction with SRY.

We have used recombinant SRY to determine the optimal DNA target sequence from a pool of random DNA sequences. The ^{32}P-labelled PCR products from alternate cycles of selection were incubated with *E.coli* extract containing SRY protein and assayed by gel retardation (Figure 1). The random oligonucleotide (0 selection cycles) failed to form detectable protein–DNA complexes (Figure 1, lane 1). By the third cycle of selection, enrichment for SRY-binding oligonucleotides was clearly evident (Figure 1, left panel). No protein–DNA complex is apparent in the control extract harbouring the expression vector lacking the *SRY* gene (Figure 1, right panel).

After nine cycles of selection, DNA was sequenced. All individual clones recovered were AT rich and AACAAT/A occurs in 13 of 29 sequences (Figure 2). We had shown previously that mutant DNA sites ACCAAA, AACCAA, AACAGA and AACAAG bound poorly to SRY (3); such sequences are not selected by SRY (Figure 2). When compared to the TCF-1 target sequence AACAAAG, no preference for G is evident at the seventh position. Also, we observe a strong preference for A or T (26 of 29) in the nucleotide preceding the site; this −1 position is partially protected by SRY in DEPC interference assays (5). Overall, a consensus DNA binding site for SRY is A/TAACAAT/A where the A is favoured in the −1 position (15 of 26) and the T is favoured in the sixth position (17 of 27). We investigated the strength of interaction between SRY and its consensus site compared with the TCF-1 target site. The sequence AACAATG competed 5-fold better for SRY (Figure 3, lanes 7,8) than AACAAAG (Figure 3, lanes 5,6), indicating that we have identified a high affinity site for SRY and have confirmed the preference for T in the sixth position.

The lack of preference in the seventh position and the preponderance of T in the sixth position suggest that SRY has a specificity that differs from TCF1/LEF1. Consistent with this idea, mouse Sry binds AACAATG better than does LEF1, although the opposite is true for AACAAAG (5). By methods similar to ours, the consensus AACAAT was derived for Sox-5, a testis-specific, Sry-related protein (9). We predict that all HMG box proteins with DNA-sequence specificity may interact with motifs similar to that found here contacting A-T rich regions present in the minor groove and inducing a dramatic bend in the

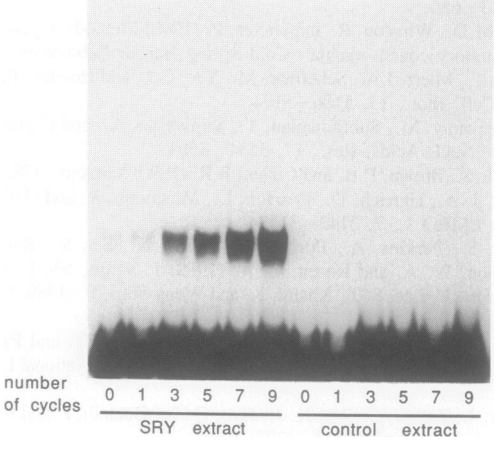

Figure 1. Electrophoretic mobility-shift analysis of ^{32}P-labelled random DNA oligonucleotides following selection by recombinant SRY. The number of cycles of selection is shown under each lane (O denotes the random oligonucleotide, 5′-CAGGTCAGTTCAGCGGATCCTGTCG(A/G/T/C)$_{26}$AGGCGAATTCA-GTGCAACTGCAGC-3′). Reactions contained 1–2 μg FLAG-SRY *E.coli* cytoplasmic extract (left panel) or control extract (right panel) and 10 000 cpm of DNA fragment. SRY production, reaction and gel-running conditions were as described previously (3). During the selection procedure, gels were dried down and exposed overnight on Kodak X-AR film for band localisation (13). Initially, the position of the complex was not visible so the HuSRY probe (5′-GTTAA-CGTAACAAAGAATCTGGTAGA) was used as a guide in adjacent tracks. The selected DNA/SRY complex was excised and the gel sliced into mm^2 pieces and a single piece was amplified by PCR in the presence of [α^{32}P]dCTP, gel purified, and 0.2 ng was selected as before (14). After nine cycles of selection the PCR product was digested with *Bam*HI and *Eco*RI and cloned into pBluescript for sequencing.

* To whom correspondence should be addressed

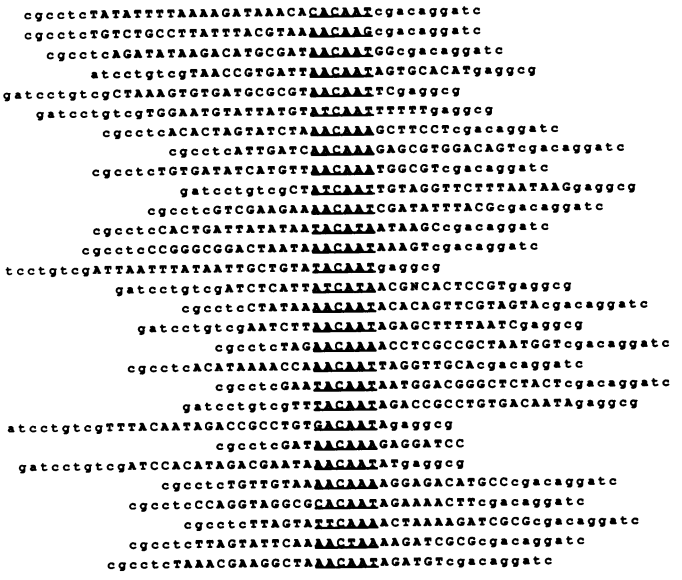

Figure 2. Alignment of selected SRY binding sites. Upper case letters represent the random region of the original random oligonucleotide probe. Sequences were aligned about their core binding sequence, A/T A A C A A T.

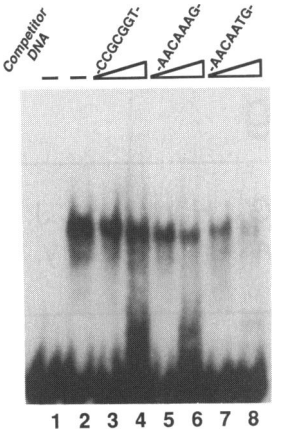

Figure 3. Competition for SRY-DNA binding. Lane 1. labelled TMUT probe (5′-GTTAACGTAACAATGAATCTGGTAGA) (100 fmol) alone. Lane 2–8, probe plus 1 μl SRY *E.coli* extract. Reactions were incubated with competitor DNA (10 pmol or 100 pmol) for 10 min prior to the addition of probe. Competitor DNA differ from TMUT probe in their core sequence as indicated.

target DNA (5,10). Also, it seems likely that this class will form subgroups, e.g. SRY/SOX and LEF1/TCF1, with unique DNA specificities.

Knowledge of a consensus DNA binding motif for SRY may alert us to regulatory sequences present in candidate downstream genes; for example the *WT1* promoter contains an AACAAT site (12) while *AMH* and *aromatase* promoters do not (7). Also, chromatin immunopurification, as for Hox-C8 (11), might identify *in vivo* targets for SRY which could be examined for the presence of AACAAT.

ACKNOWLEDGEMENTS

We thank Roy Pollock and Richard Treisman for helpful advice. This work was supported by the MRC (UK) and the Wellcome Trust.

REFERENCES

1. Landsman,D. and Bustin,M. (1993) *Bioessays* **15**, 539–546.
2. Ner,S.S. (1992) *Current Biology* **2**, 208–210.
3. Harley,V.R., Jackson,D.I., Hextall,P.J., Hawkins,J.R. *et al.* (1992) *Science* **255**, 453–456.
4. van de Wetering,M. and Clevers,H. (1992) *EMBO J.* **11**, 3039–3044.
5. Giese,K., Cox,J. and Grosschedl,R. (1992) *Cell* **69**, 1–20.
6. Nasrin,N., Buggs,C., Kong,X.F., Carnazza,J. *et al.* (1991) *Nature* **354**, 317–320.
7. Haaq,C.M., King,C.-Y., Donahoe,P.K. and Weiss,M.A. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 1097–1101.
8. Harley,V.R. and Goodfellow,P.N. (in press) *Mol. Rep. Dev.*
9. Denny,P., Swift,S., Connor,F. and Ashworth,A. (1992) *EMBO J.* **11**, 3705–3712.
10. Ferrari,S., Harley,V.R., Pontiggia,A., Lovell-Badge,R. *et al.* (1992) *EMBO J.* **11**, 4497–4506.
11. Tomotsune,D., Shoji,H., Wakamatsu,Y., Kondoh,H. and Takahasi,N. (1993) *Nature* **365**, 69–72.
12. Pelletier,J., Schalling,M., Buckler,A.J., Rogers,A., Habre,D.A. and Houseman,D. (1991) *Genes Dev.* **5**, 1345–1356.
13. Mavrothalassitis,G, Beal,G. and Papas,T. (1990) *DNA Cell Biol.* **9**, 783–788.
14. Pollock,R. and Treisman,R. (1990) *Nucleic Acids. Res.* **18**, 6197–6204.