DNA binding and bending properties of the post-meiotically expressed Sry-related protein Sox-5

Frances Connor, Peter D.Cary[1], Christopher M.Read[1], Nicola S.Preston[1], Paul C.Driscoll[2], Paul Denny[+], Colyn Crane-Robinson[1] and Alan Ashworth*

CRC Centre for Cell and Molecular Biology, Chester Beatty Laboratories, The Institute of Cancer Research, Fulham Road, London SW3 6JB, [1]Biophysics Laboratories, University of Portsmouth, Portsmouth PO1 2DT and [2]Department of Biochemistry and OCMS, University of Oxford, Oxford OX1 3QU, UK

ABSTRACT

Sox-5 is one of a family of genes which show homology to the HMG box region of the testis determining gene *SRY*. We have used indirect immunofluorescence to show that Sox-5 protein is localized to the nucleus of post-meiotic round spermatids in the mouse testis. *In vitro* footprinting and gel retardation assays demonstrate that Sox-5 binds specifically to the sequence AACAAT with moderately high affinity (K_d of ~10^{-9} M). Moreover, interaction of Sox-5 with its target DNA induces a significant bend in the DNA, characteristic of HMG box proteins. Circular dichroism spectroscopy of the Sox-5 HMG box and its specific complex with DNA shows an alteration in the DNA spectrum, perhaps as a consequence of DNA bending, but none in the protein spectrum on complex formation. The dependence of the change in the CD spectrum with protein to DNA ratio demonstrates the formation of a 1:1 complex. Analysis of the structure of the Sox-5 HMG box by 2D NMR suggests that both the location of helical secondary structure as well as the tertiary structure is similar to that of HMG1 box 2.

INTRODUCTION

The development of mammals as males depends upon the presence of a dominantly acting gene, the testis determining gene, on the Y chromosome. This gene directs the divergence during embryogenesis of the indifferent gonad from the 'default' ovarian pathway to that of the testis (1). There is now overwhelming evidence that *Sry* encodes this gene, culminating in the demonstration of testis formation in female mice transgenic for *Sry* (2). A distinctive feature of Sry is the presence of a structural protein motif, the HMG box, which was initially identified as a region of homology with the high mobility group (HMG) non-histone chromosomal proteins (3–5). A growing family of diverse proteins which share this motif have been identified. Two subfamilies have been defined: essentially non-sequence-specific DNA binding proteins, often with multiple HMG boxes, and proteins with a single sequence-specific HMG box (6). Sry falls into the latter group, together with the putative lymphoid differentiation regulators TCF1 (7) and LEF-1 (8) and the fungal mating type proteins Mat-Mc (9) and Mt-a1 (10).

On the basis of homology (>60%) between their HMG boxes, a family of *Sry*-related genes have been described in mammals as well as birds, reptiles, amphibians and insects (5,11,12). These have been named *Sox* ('Sry box') genes. The function of these *Sox* genes is at present unknown but roles in neurogenesis for *Sox-1* to *Sox-3* have been proposed (13). We have previously described an *Sry*-related gene *Sox-5* which is expressed exclusively in post-meiotic cells of the adult mouse testis (14). Sox-5 protein binds DNA sequence-specifically *in vitro* and we have suggested it has a role in regulating gene expression during the post-meiotic phase of spermatogenesis. Here we demonstrate that Sox-5 is located in the nucleus of spermatogenic cells and investigate the structure and DNA binding properties of its HMG box. We show that in common with Sry, Sox-5 induces a large bend in the double helix upon DNA binding, due primarily to the presence of the HMG box. Furthermore CD and NMR studies establish that the HMG box of Sox-5 forms a 1:1 complex with its DNA recognition sequence and has a fold closely related to that established for HMG1 box 2.

MATERIALS AND METHODS

Germ cell preparation and immunofluorescence

Testes from adult MF1 mice were dissected and a suspension of germ cells prepared from the seminiferous tubules according to Willison *et al.* (15). Germ cell suspensions (~3–6 ×10^6 cells/ml) were fixed in 3% paraformaldehyde in PBS for 10 min at room temperature and then quenched in 50 mM NH₄Cl/PBS.

*To whom correspondence should be addressed

[+]Present address: Nuffield Department of Surgery, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK

Fixed cells were allowed to settle onto poly-L-lysine (100 mg/ml) coated coverslips for 60 min before being permeabilised by incubation for 1 min in 100% methanol at −20°C. Cells were subsequently blocked in PBS plus 0.2% gelatin (10 min) and stained with the first antibody diluted in PBS plus 0.2% gelatin for 30 min. Affinity-purified anti-Sox-5 peptide antibody (14) was used. Control immunofluorescence used the same antibody pre-competed overnight at 4°C with a 500× molar excess of either the Sox-5 peptide or an irrelevant peptide. Following extensive washing in PBS and blocking in PBS plus 0.2% gelatin, bound antibodies were visualised using fluorescein-conjugated goat anti-rabbit antibody (Pierce). Thoroughly washed coverslips were mounted in anti-quench mountant (Vectashield, Vector Laboratories) or Moviol containing 0.1% p-phenylene diamine (Sigma). Images were recorded using a Bio-Rad MRC600 confocal microscope.

Production of recombinant Sox-5 and Sry proteins

Full length Sox-5 and murine Sry were expressed as glutathione-S-transferase (GST) fusion proteins in bacteria as described (14). Nucleotides 645−881 from the Sox-5 cDNA clone pD588 (14), encompassing the HMG box, were amplified by PCR and subcloned into the pGEX-2T expression vector (16). The GST fusion protein was expressed in BL21(DE3) pLysS host bacteria (17) and purified using glutathione−Sepharose beads (16). After cleavage of the fusion protein with bovine plasma thrombin (7.5 units/l culture), the released Sox-5 HMG box was purified either by a combination of ion-exchange and gel filtration chromatography (18) or by semi-preparative C4 reverse-phase HPLC chromatography. The ability of the protein to refold correctly after reverse-phase HPLC was checked by comparison of NMR spectra with those of a sample that had been purified only by ion-exchange/gel filtration and so never denatured.

Electro-spray mass spectrometry of the purified Sox-5 HMG box gave a molecular mass of 9805.21 ± 0.56 daltons in accordance with that expected for the expressed 79 amino acids and the two additional vector-encoded N-terminal residues (Gly−Ser). The 79 residues of the recombinant Sox-5 HMG box precisely correspond in position to the 79 residue HMG1 box 2 studied in Read et al. (18) (see Figure 5).

DNase I protection assays

DNAs for DNase I footprinting were derived from clone p4691 (14) containing the in vitro Sox-5 binding site (CCCCCCTAA-AAG\underline{AACAA}TACTTCCCA). The NotI linearised plasmid was labelled using a Klenow DNA polymerase I fill-in reaction with [α-^{32}P]dCTP. Following XhoI digestion the NotI−XhoI 132 bp restriction fragment, labelled on the 'top' strand only, was purified over polyacrylamide. A 147 bp XhoI−SacI restriction fragment labelled on the 'bottom' strand only was made in a similar way.

Approximately 20−40 fmol of probe DNA and 0−5 μg of GST−Sox-5, Sox-5 HMG box or GST−Sry were mixed together in the presence of 0.1−1.0 μg poly (dI−dC) in footprint binding buffer (20 mM HEPES (pH 7.9), 50 mM NaCl, 5 mM MgCl$_2$, 0.1 mM EDTA, 1 mM DTT and 20% glycerol) to a final volume of 100 μl. After 45 min incubation on ice, the samples were treated with 1−10 units DNase I in 1 mM CaCl$_2$ for 30 s at room temprature. 100 μl of Stop solution (1% SDS, 100 mM Tris−HCl, pH 8.0, 10 mM EDTA, 0.4 mg/ml Proteinase K and 0.1 mg/ml glycogen) were added and the samples incubated at 37°C for 20 min then at 70°C for 2 min. DNA samples were

phenol/chloroform extracted, ethanol precipitated and separated by denaturing polyacrylamide gel electrophoresis followed by autoradiography.

Equilibrium binding constant determination

The equilibrium binding constants of full length Sox-5 and HMG box Sox-5 proteins were measured by a serial dilution experiment in which the concentrations of protein added to the DNA duplex were varied without altering their ratio (19). The duplex was made by 5'-end labelling the oligonucleotide BS5 using T4 polynucleotide kinase and [γ-^{32}P]ATP, then annealing the unlabelled complementary strand BS6. The sequences of the oligonucleotides (with the in vitro Sox-5 binding site underlined) are: BS5 5'-CTAGCACTAT\underline{AACAA}TACAAGCCGCGG-3' and BS6 5'-CTAGCCGCGGCTTGT\underline{ATTGTT}ATAGTG-3'.

A typical dilution experiment was as follows. The initial reaction mixture contained 10 nM BS5+6 duplex and 100 nM Sox-5 protein in 36 μl of EMSA binding buffer (10 mM HEPES, pH 7.9, 60 mM KCl, 1 mM DTT, 1 mM EDTA, 0.33 mg/ml BSA and 12% glycerol). Complexes were allowed to equilibrate for 30 min at room temperature. 1.5-Fold serial dilutions were carried out in the binding buffer. Following equilibration at room temperature, samples were electrophoresed on a non-denaturing 4% polyacrylamide gel (0.5× TBE; 40:1 acrylamide:bis) at room temperature using 10 V/cm for 90 min. The gels were dried and the radioactivity in free and bound DNA measured on a phosphorimager (Molecular Dynamics). The binding constant was determined according to Liu-Johnson et al. (19).

Circular permutation assay

Annealed synthetic oligonucleotides containing the in vitro Sox-5 binding site (BS5+6) were cloned into the XbaI site of the circular permutation vector pBend2 (20). Circularly permuted DNA fragments were made by cleavage with one of the restriction enzymes indicated in Figure 3A and purified by agarose gel electrophoresis. The DNAs were labelled either by fill-in using the Klenow fragment of DNA polymerase I and [α-^{32}P]dCTP or, following dephosphorylation, with T4 polynucleotide kinase and [γ-^{32}P]ATP.

EMSAs were performed essentially as described (14). Briefly, 25−100 ng of purified protein were pre-incubated in 15 μl of EMSA binding buffer with 25−500 ng poly (dI−dC) for 5 min at room temperature, then ~10,000 cpm of DNA was added and incubated for a further 20 min at room temperature. The reactions were electrophoresed on a non-denaturing 8% polyacrylamide gel (0.5× TBE; 75:1 acrylamide:bis) at 10 V/cm at room temperature for 4−5 h. Gels were dried prior to autoradiography. Bend parameters were calculated according to Thompson and Landy (21).

Circular dichroism

The two oligonucleotides 16A and B (5'-ACTAT\underline{AACAAT}AC-AAG-3' and 5'-CTTGT\underline{ATTGTT}ATAGT-3' respectively, with the in vitro Sox-5 binding site underlined) were mixed in equal molar amounts (as determined from their UV extinction coefficients) in 10 mM Tris−HCl (pH 7.5), 50 mM NaCl, 10 mM MgCl$_2$ and then incubated at 70°C for 10 min, followed by slow cooling to 16°C. Duplex DNA was separated from residual oligonucleotides by gel filtration chromatography on a Superose 12HR 10/30 column (Pharmacia) equilibrated in 1 mM Tris−HCl (pH 6.5), 20 mM NaCl, 2 mM MgCl$_2$. The peak

containing duplex DNA (as judged by polyacrylamide gel electrophoresis) was used directly for CD measurements.

Spectra were obtained on a Jasco A20 spectropolarimeter with purging of nitrogen gas, at 20°C. The instrument was calibrated for ellipticity and wavelength against a standard of camphor sulphonic acid. CD spectra were recorded in 20 mM potassium phosphate (pH 6.0) buffer between 330 and 186 nm, using far UV quartz cells of pathlength 0.12, 1.0 and 5.0 mm.

NMR spectrometry

Initial NMR spectra were recorded in one-dimension on a Jeol 270 MHz instrument at 20°C with samples dissolved in 100% D_2O. Two-dimensional proton NMR spectra were recorded on Bruker and General Electric 600 MHz instruments (OCMS, University of Oxford) in 90% H_2O/10% D_2O and 100% D_2O at 20°C, at a protein concentration of 3.0 mM in 75 mM potassium phosphate buffer (pH 6.2; uncorrected glass electrode reading), 0.5 mM DTT. 2D ^1H nuclear Overhauser enhancement (NOESY) and 2D ^1H homonuclear Hartmann—Hahn (HOHAHA) spectra were obtained as described previously (18). Data were processed using FELIX v2.30 (BioSym) running on a Silicon Graphics Indigo 2 XZ workstation.

RESULTS AND DISCUSSION

Sox-5 is located in the nucleus of round spermatids

We wished to investigate whether Sox-5, a potential transcription factor expressed in the testis, was located in the nucleus of the cells in which it is expressed. The main male germ cell types are the diploid pre-meiotic spermatogonia, the meiotic spermatocytes and the post-meiotic haploid cells, which are the round and elongating spermatids and mature spermatozoa (22). We have previously shown using immunoblot analysis that Sox-5 is restricted to post-meiotic germ cells, being found at highest levels in round spermatids (14). Here we have used indirect immunofluorescence to determine the subcellular location of endogenous Sox-5 in adult male murine germ cells.

Representative cells from all stages of spermatogenesis were prepared from adult mice testes and stained with anti-Sox-5 antibody. Sox-5 was found in cells which were identified as round
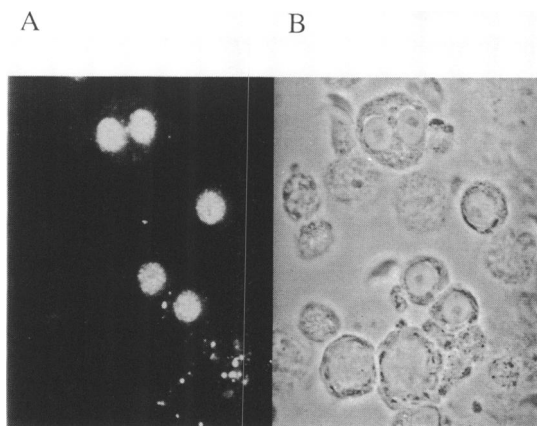

Figure 1. Sox-5 is located in the nucleus of round spermatids. Adult male murine germ cells were fixed, permeabilised and incubated with affinity-purified anti-Sox-5 peptide antibody followed by fluorescein-labelled anti-rabbit antibody. Strong staining was seen in the nuclei of round spermatids (**A**). The corresponding phase contrast image is shown in (**B**).

spermatids, as judged by their size and shape. Pre-meiotic diploid cells did not stain (Figure 1A and B). The fluorescence could be eliminated by pre-competing the antibody with a Sox-5 peptide but not with an irrelevant peptide (data not shown). These observations confirm the round spermatids as the main Sox-5 expressing cells. This analysis also demonstrates that Sox-5 is predominantly localised to the nucleus, as might be expected for a potential transcription factor. The same result was obtained by Western blot analysis of cytoplasmic and nuclear fractions of germ cells (data not shown). The observed fluorescence of the acrosomal material in late spermatids and mature spermatozoa is artefactual. It is also seen in controls using either the fluorescein-labelled second antibody alone or using the anti-Sox-5 antibody pre-competed with an excess of Sox-5 peptide (data not shown). This phenomenon has been noted with other antibodies (K.Willison, personnal communication). No other significant background staining was observed.

Sox-5 HMG box binds DNA sequence specifically

We have previously shown that Sox-5 binds preferentially to oligonucleotides containing the sequence motif AACAAT in an *in vitro* DNA binding site selection assay (14). A DNase I protection assay was used to locate the Sox-5 binding region to the area of protein—DNA contact. DNase I digestion ladders were obtained for a DNA restriction fragment derived from one of the clones containing a Sox-5 selected binding sequence (Figure 2). Comparison of control digestions containing no Sox-5 protein with those containing an excess of recombinant Sox-5 HMG box protein reveals protected 'footprints' of 14—15 nucleotides on the 'top' strand, i.e. that containing the AACAAT sequence, and of 13 nucleotides on the opposite strand. The Sox-5 selected AACAAT motif is contained within the protected regions in both orientations. The GST—full length Sox-5 fusion protein shows essentially the same pattern of DNase I protection, though the bottom strand footprint is slightly larger.

As for Sox-5, the *in vivo* targets of SRY are unknown. Initial investigations identified DNA binding sites for SRY derived from known binding motifs of other HMG box proteins (14,23—25). Site selection assays have defined the sequence AACAAT as a high affinity binding site for SRY *in vitro* (26). In our DNase I protection assay recombinant mouse GST—Sry fusion protein shows 'footprints' encompassing the AACAAT motif (Figure 2). The protected regions are similar to those seen with the Sox-5 HMG box. However, recent findings indicate that DNA recognition by SRY may be species-specific; mouse and human SRY HMG boxes show distinct DNA binding and bending properties (27).

The equilibrium binding constant (K_d) of Sox-5 HMG box to its *in vitro* recognition motif was estimated from a serial dilution experiment in which the total concentration of added protein to DNA duplex was varied without altering the ratio. Data were plotted according to the method described by Liu-Johnson *et al.* (19). From several such experiments we estimate a K_d of ~2 $\times 10^{-9}$ M, suggesting a moderately high affinity of specific DNA binding by the Sox-5 HMG box. This compares to values of 3 $\times 10^{-9}$ M and 50 $\times 10^{-9}$ M for mouse and human SRY respectively (27). Our estimate of the binding affinity of Sox-5 is also comparable to those determined for the binding of LEF-1 and TCF1 to their recognition motifs (28, unpublished data in 29). Interestingly the binding affinity of Sox-4, a 'classical' transcription factor capable of transactivation, is approximately 30-fold greater (29).

The HMG box of Sox-5 bends DNA

A structural role for the HMG domain in the recognition and production of bends in DNA has been suggested (30). We used a circular permutation assay to show that the Sox-5 HMG box, which binds DNA sequence-specifically, is similarly able to bend DNA. The AACAAT motif was cloned into the vector pBend2 (20). Digestion of the resulting plasmid with a number of restriction endonucleases gave a set of probe DNAs of almost identical length and base composition but with the AACAAT motif at different positions (Figure 3A). Wu and Crothers (31) observed that the electrophoretic mobility of bent DNA is dependent upon the location of the bend; reduction of migration is greatest when the DNA is bent at its centre and least when near its end. Recombinant Sox-5 HMG box forms complexes with the circularly permuted DNA probes (Figure 3B). Their mobility is clearly dependent upon the position of the binding motif in the fragment. The free DNA duplexes have virtually identical mobilities, showing that they are not intrinsically distorted.

The mobility of DNA in a gel has not been quantitatively described in full. However, this assay allows an empirical estimation of the bend locus and magnitude. The centre of bending induced by the Sox-5 HMG box maps to the centre of the AACAAT motif (Figure 3C). The size of the protein-induced bend in a DNA fragment can be estimated using the empirical equation proposed by Thompson and Landy (21). This involves a comparison of the ratio of mobilities of molecules that have a bend of unknown magnitude in the centre or at the end to the same ratio for molecules with standard bends. Using this approach, the angle by which the DNA is bent from linearity by the binding of Sox-5 HMG box is estimated at ~74°. However, the circular permutation analysis is sensitive not only to protein-induced bends, but also to other distortions in DNA structure caused, for example, by protein shape or DNA flexibility. These values are therefore better described as DNA flexure angles (32). The same experiment was repeated using recombinant GST−full length Sox-5 protein. The relative mobilities of the various protein−DNA complexes showed the pattern typical for a bend induced by protein; the site of flexure localises to the AACAAT motif and the estimated DNA flexure angle is 113° (data not shown).

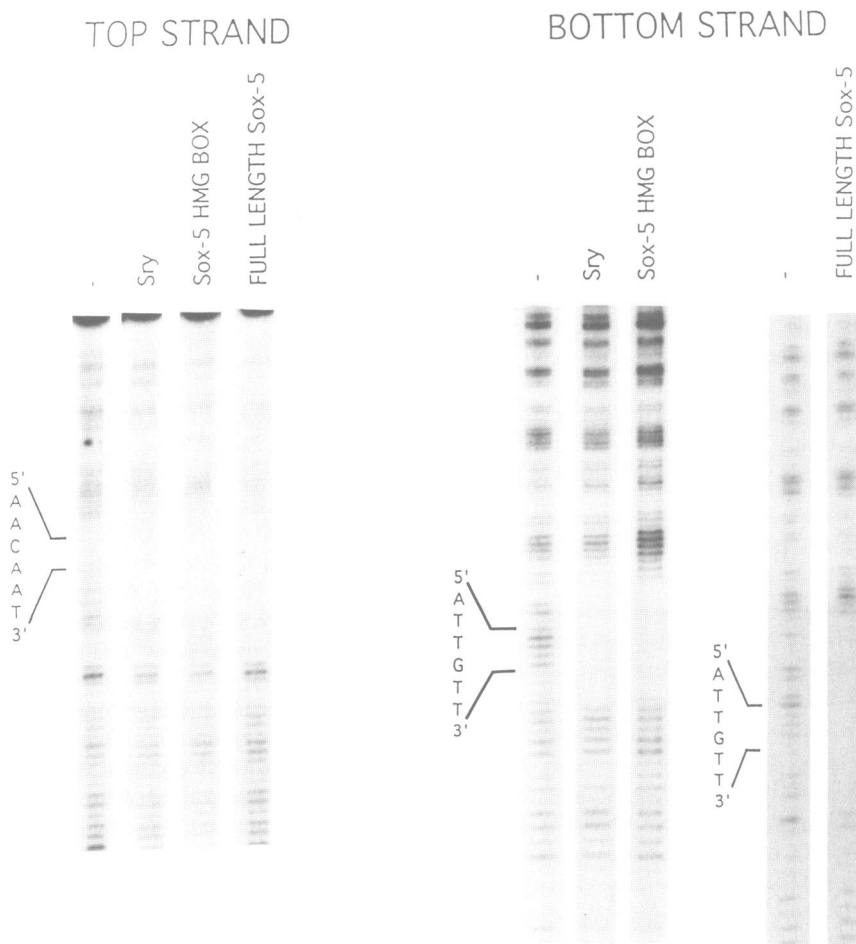Estimations of the angle of bending induced by mouse Sry binding to AACAAT (25) and by human SRY to AACAAG (33)

Figure 2. Sequence-specific DNA binding of recombinant Sox-5 and Sry. DNase I protection analysis is shown for GST−Sry, Sox-5 HMG box and GST−Sox-5 on DNA containing the Sox-5 binding motif, AACAAT. Restriction fragments were ^{32}P-labelled on either the 'TOP' or 'BOTTOM' strand and incubated with or without protein. The subsequent products of DNase I digestion were analysed on denaturing polyacrylamide gels. Sequencing reactions were run in adjacent tracks as size markers.
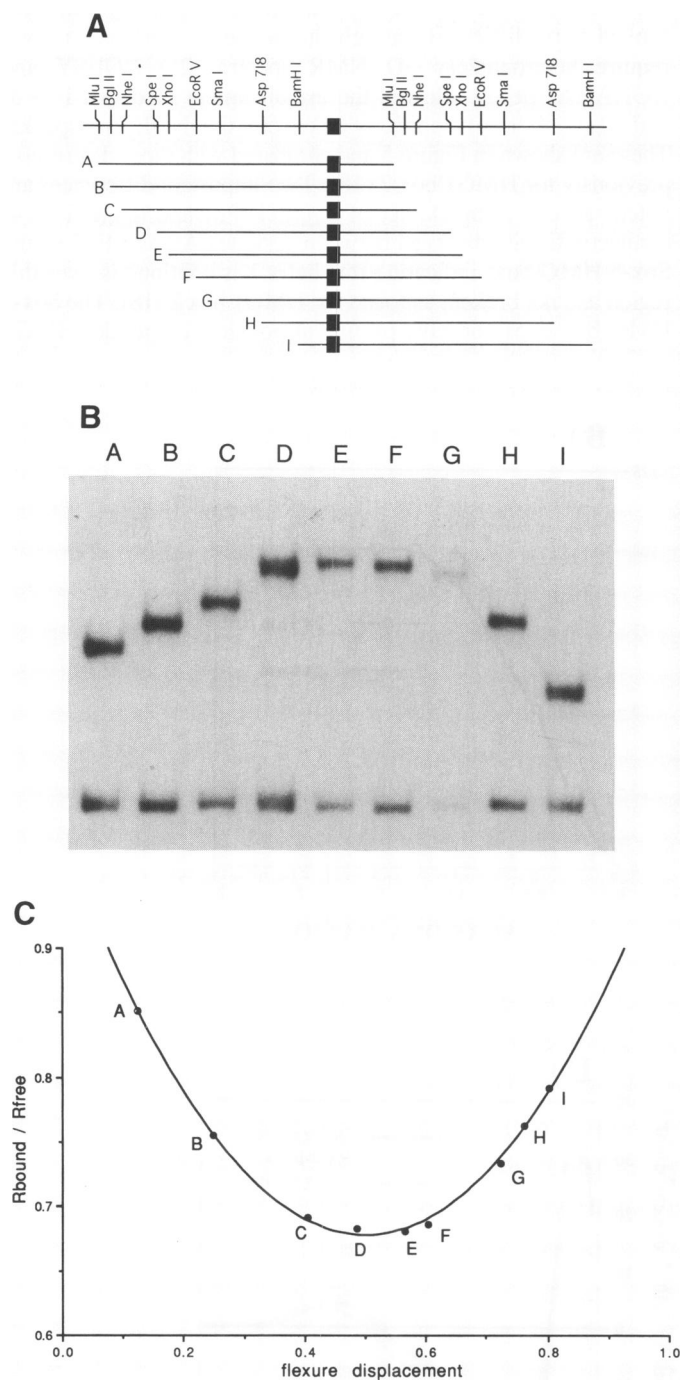
Figure 3. DNA bending by Sox-5 analysed by circular permutation assay. (A) Structure of DNA probes used for the bending analysis. The Sox-5 *in vitro* DNA binding site (filled box) was cloned between the two direct repeats of vector pBend2 (20). Cleavage of the resulting plasmid at the restriction sites shown effectively permutes the binding site along a DNA fragment, maintaining the length (148 or 152 bp) and nucleotide sequence of the probes (A−I). The restriction enzymes used were *Mlu*I (A), *Bgl*II (B), *Nhe*I (C), *Spe*I (D), *Xho*I (E), *Eco*RV (F), *Sma*I (G), *Asp*718 (H) and *Bam*HI (I). (B) EMSA of Sox-5 HMG box complexes with the circularly permuted DNA probes (A−I). (C) Mapping of the centre of bending to the *in vitro* Sox-5 binding site. The mobilities of the protein−DNA complexes (R_{bound}), normalised to those of the free probe (R_{free}), were plotted against the flexure displacement of the probes. The flexure displacement is defined, according to Ferrari *et al*. (33), as the distance of the centre of the Sox-5 *in vitro* binding motif from the 5' end of the probe divided by the total length of the probe. The minimum locates the centre of bending to the AACAAT motif.

have been reported as 85° and 83°, respectively. In agreement with this we calculate a deviation from linearity of ~73°−90° by binding of recombinant mouse GST−Sry to the set of probes shown in Figure 4A (data not shown). The validity of the bend angles calculated must be tempered by the fact that the nature of the protein−DNA interaction is as yet unknown and that nucleotides flanking the core AACAAT sequence may influence the angle of bend (27). Nevertheless, these experiments demonstrate that the Sox-5 HMG box, either alone or in the context of the full length protein, can induce a strong bend in DNA at its target binding site.

DNA binding of Sox-5 assayed using circular dichroism indicates a 1:1 stoichiometry, DNA distortion but no change in protein secondary structure

The CD spectrum of B-form DNA exhibits a positive maximum at ~275 nm where there is no ellipticity from the intrinsic Cotton effect of protein and only very weak extrinsic Cotton effects. Conformational changes in the DNA can thus be monitored in this region. Titration of the Sox-5 HMG box onto the 16mer duplex containing the recognition sequence led to a blue shift of the maximum to 265 nm, without significant change in the peak ellipticity (Figure 4A). This allowed quantitative monitoring of binding by measuring the decrease in ellipticity at 285 nm and the increase at 260 nm, plotted in Figure 4B as a percentage of the maximum change. It is clear that the ellipticity changes rise to a plateau at close to a 1:1 molar ratio of Sox-5 HMG box:16mer DNA, so indicating the formation of a bi-molecular complex. This accords with the finding that the sequence-specific HMG box LEF-1 binds to its recognition site as a monomer (28).

The blue shift of the 275 nm maximum cannot be readily interpreted in structural terms but is probably due to the bending of the DNA induced by protein binding. It is a quite different effect from the change induced by histone binding in the nucleosome to produce an approximately continuous DNA bending which results in a reduction of the ellipticity at 275 nm without significant shift of the maximum (34). It might therefore be that the blue shift is a consequence of DNA kinking or a significant unwinding of the duplex, but comparison with other HMG boxes will be required to establish if this feature is genuinely a characteristic of the induced DNA distortion.

The CD spectrum of the Sox-5 HMG box alone (Figure 4C) shows the characteristic features of an α-helical protein: strong negative maxima at 222 and 207 nm, together with a strong positive maximum at 190 nm. Secondary structure calculation by the method of Compton and Johnson (35) gave a best fit with 55% α-helix, 20% turns and 25% irregular, without the need to introduce a β-sheet component. Since the protein contains 81 amino acids, this implies 44 or 45 α-helical residues.

The complete CD spectrum of a 1:1 complex between the Sox-5 HMG box and the 16mer duplex is shown in Figure 4D, together with a curve representing the sum of the spectra of the protein and the DNA taken separately. The blue shift in the 275 nm DNA maximum is readily apparent, however there is no evidence of any differences in the CD spectrum at lower wavelengths, a region predominantly due to the Sox-5 HMG box protein signal. This indicates that there are no significant changes in the secondary structure of the protein on binding to DNA. However, this does not rule out any macroscopic changes in the tertiary structure or 'shape' of the fold, such as have been proposed in an induced fit model of HMG box binding (33). In contrast, in a study of the binding of the TCF-1 HMG box to

its recognition sequence, van Houte *et al.* (36) recorded an apparent drop in α-helix content of the protein on DNA binding, as seen from an apparent reduction in the negative ellipticity at 222 nm from −10,000 to −6500° cm² dmol⁻¹. Since in the present work the negative ellipticity at 222 nm for the Sox-5 HMG box is significantly greater (at −17,000° cm² dmol⁻¹), it is difficult to compare the two results.

The location of secondary structure in the Sox-5 HMG box is similar, but not identical, to that in HMG1 box 2

The majority of the amino acid spin systems have been assigned from 2D NMR spectra of the Sox-5 HMG box. Certain segments

within the last third of the molecule (helix 3 and beyond) have proved very difficult to assign due to peak overlaps and will require heteronuclear 3D NMR spectra. 2D NOESY and HOHAHA spectra allowed the establishment of $d_{\alpha N}(i, i + 3)$ and $d_{\alpha N}(i, i + 4)$ connectivities from the NOE crosspeaks. These are shown in Figure 5 and compared with those determined previously for HMG1 box 2 (18). Two important differences are noted: a very complete set of $d_{\alpha N}(i, i +3)$ and $d_{\alpha N}(i, i + 4)$ connectivities is observed between residues A9 and F25 of the Sox-5 HMG box, indicating that helix 1 is continuous over this region and not broken, as found for HMG1 box 2 (18). The Sox-5 HMG box does not have proline at position 19 (unlike HMG1
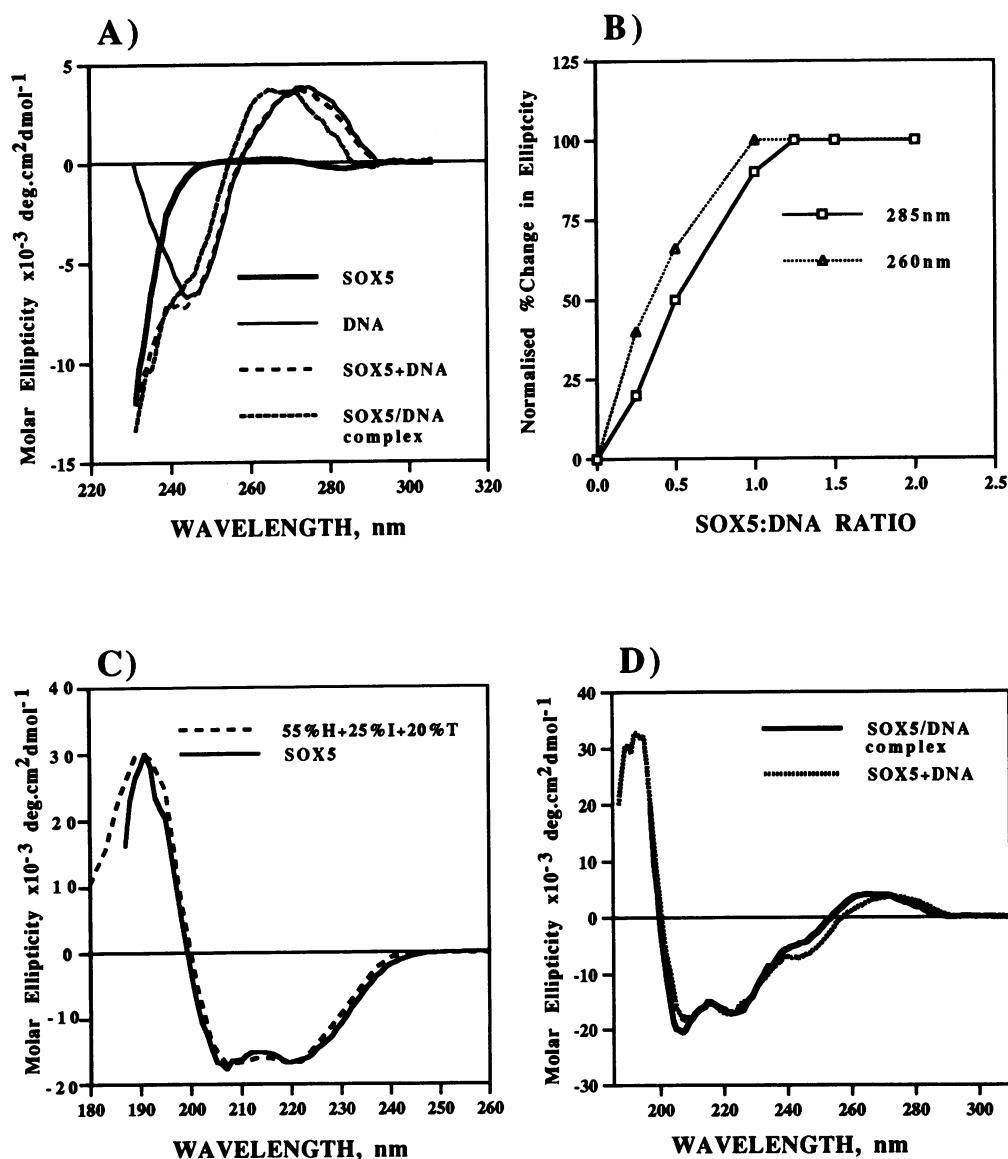

Figure 4. Circular dichroism measurements of Sox-5 and its DNA binding site. (A) CD spectra in the 230−310 nm region of the Sox-5 HMG box, the 16mer DNA duplex containing the AACAAT recognition sequence and a 1:1 complex formed between Sox-5 HMG box and the 16mer DNA. Also shown is the addition of the Sox-5 HMG box and 16mer DNA CD spectra. (B) Plot of the percentage decrease in ellipticity at 285 nm and the increase at 260 nm, normalised to the maximum change, at various Sox-5 HMG box to 16mer DNA ratios. (C) CD spectrum of the Sox-5 HMG box and the curve-fitted spectrum corresponding to 55% α-helix, 25% irregular and 20% turns. (D) CD spectrum of the 1:1 complex of Sox-5 HMG box and 16mer DNA. Also shown is the addition of the CD spectra of the Sox-5 HMG box and 16mer DNA.

box 2) and this could be one of the determining features of the continuous helix. In their model of HMG1 box 2, Weir *et al.* (37) show a continuous helix 1, albeit somewhat bent around

Table 1. Comparison of selected NOE through space contacts in the SOX-5 HMG box with those of HMG1 box2

SOX-5 HMG box			HMG1 box2		
Proton(s)	Proton(s)	NOE Intensity[†]	Proton(s)	Proton(s)	NOE Intensity[†]
Ala9 HB*	Trp41 HZ2	S	Ala9 HB*	Trp41 HZ2	S
Phe10 HD*	Trp41 HB*	M	Phe10 HD*	Trp41 HB*	S/M
Val12 HG*	Tyr52 HD*	M	Leu12 HD*	Tyr52 HD*	M
Trp13 HH2	Arg40 HB*	M	Phe13 HE*	Met40 HB*	M
His29 HN	Asn32 HB*	M	Ser29 HN	Asp32 HB*	M
Met44 HE*	Tyr52 HE*	S	Thr44 HG2*	Tyr52 HE*	M
Lys49 HG1	Trp41 HZ2	M	Lys49 HG1	Trp41 HZ2	M
Lys49 HA	Tyr52 HD*	S	Lys49 HA	Tyr52 HD*	S

[†]NOE intensity classes represent the following distance classes between protons: S = 1.8 to 2.5 Å, S/M = 1.8 to 3.0 Å and M = 1.8 to 3.5 Å.

residue 19. Whether the presence of a discontinuity in helix 1 is a characteristic feature of non-sequence specific HMG boxes will require the determination of more structures. The second difference is the length of helix 3, which runs from Y52 to Y67 in the Sox-5 HMG box, but from P51 to Y71 in HMG1 box 2. Although there are at present some difficulties in assignment of this region, it is clear that helix 3 of the Sox-5 HMG box ends at least 4 residues before that of HMG1 box 2. This is not unexpected since residue 68 is proline in the Sox-5 HMG box, a residue conserved in a number of sequence-specific HMG boxes such as LEF-1 and the Sry/Sox family (38) but is alanine in HMG1 box 2.

Essentially common features between the boxes include: helix 2 (N32−A43 in Sox-5 HMG box and G31−N42 in HMG1 box 2) and the reverse turn between residues 25 and 28 (FPDM in the Sox-5 HMG box and HPGL in HMG1 box 2) that includes the conserved proline at position 26 and separates helices 1 and 2. The total number of helical residues determined from the NMR spectra of the Sox-5 HMG box is thus 45 (17 + 12 + 16 for
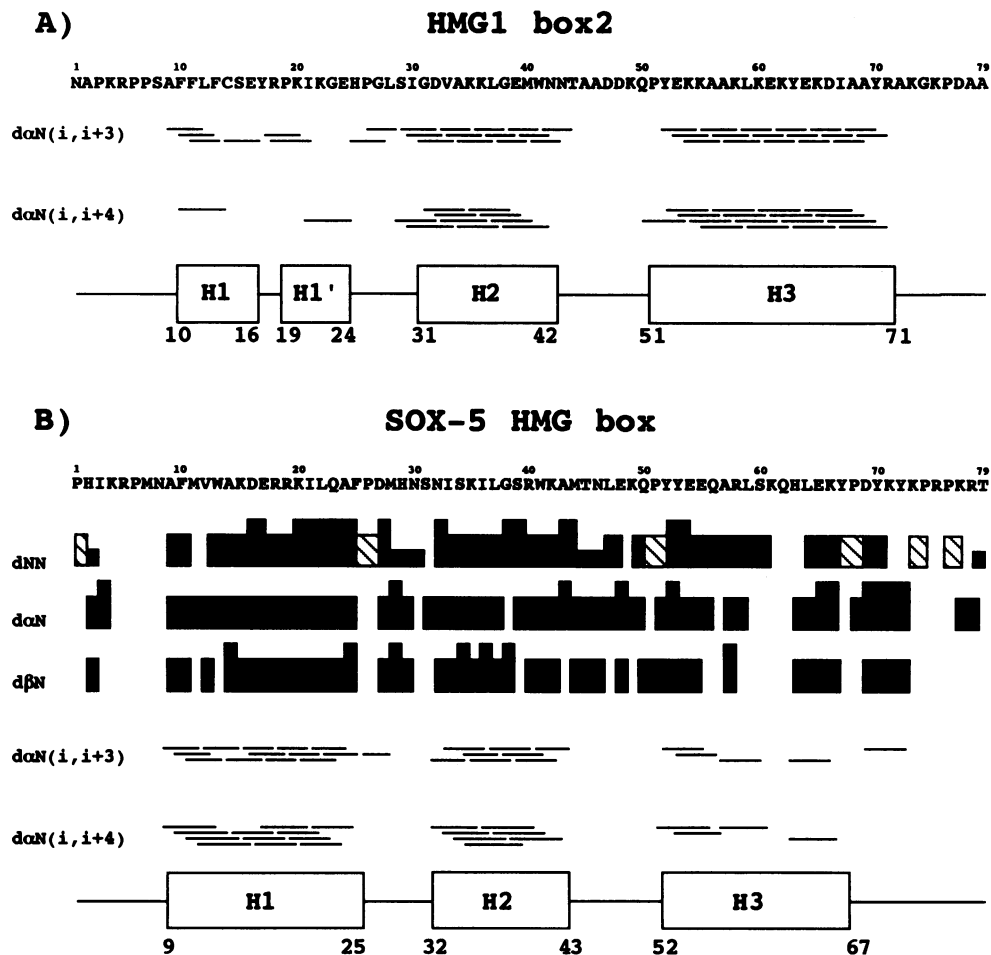

Figure 5. Comparison of selected short and medium range NOE connectivities in HMG1 box 2 and Sox-5 HMG box. (A) The top line shows the amino acid sequence of HMG1 box 2. Medium range $d_{\alpha N}$(i, i + 3) and $d_{\alpha N}$(i, i + 4) NOE connectivities are represented by lines between the two residues. Also shown are the location of the four helices identified in Read *et al.* (18). (B) The amino acid sequence of the Sox-5 HMG box is shown aligned above the sequential NOE connectivities d_{NN}, $d_{\alpha N}$, $d_{\beta N}$. The height of the solid bars indicates the NOE intensity (3 classes); for connectivities involving a proline residue, the $C_\gamma H$ protons have been used in place of the NH proton and are shown as hatched bars. Medium range $d_{\alpha N}$(i, i + 3) and $d_{\alpha N}$(i, i + 4) NOE connectivities are represented by lines between the two residues. Blank spaces represent NOE contacts that are either not observed or not completely assigned. Also shown are the locations of the three helices identified in the present work.

the 3 helices), in good accord with the CD estimate of 44/45 residues.

Tertiary structure elements are similar in the Sox-5 HMG box and HMG1 box 2

With the residue assignments at present to hand, the 2D NOESY NMR spectra have allowed a number of contacts to be established. Whilst these are not sufficient to allow a full tertiary structure determination, several long and medium range distances permit the immediate conclusion that the fold of the Sox-5 HMG box does not differ greatly from that of HMG1 box 2 (18). Table 1 lists some of the Sox-5 HMG box through-space NOE contacts compared to those observed in HMG1 box 2. For example: juxtaposition of the N-terminal part of helix 1 to the C-terminal part of helix 2 is seen from the 9/41, 10/41 and 13/40 NOE contacts, whilst the link to the N-terminal part of helix 3 is seen in the 12/52 NOE contact. The position of the turn between helices 1 and 2 relative to helix 2 is seen to be conserved from the 29/32 NOE contact. The similar positioning of the C-terminal end of helix 2 with respect to the N-terminal part of helix 3 (which have a loop of 9 residues between them), is seen from the strong 44/52 NOE contact. The conserved K49 residue in this loop, the side chain of which extends across and inside the hydrophobic core in HMG1 box 2, is seen to be in a similar position in Sox-5 by its NOE contact with W41 in helix 2. Before the start of helix 3 is an almost identical quintet of residues (DKQPY in HMG1 box 2 and EKQPY in Sox-5 HMG box at positions 48−52) that have a similar set of NOEs in both proteins, e.g. between K49 and Y52. In HMG1 box 2 this section of the chain is in the form of a helical turn that lies at an angle to, and is not continuous with helix 3. We thus expect the conformation of Sox-5 to be similar to that of HMG1 box 2 at this point.

The role of Sox-5 *in vivo* is still unclear but it seems likely to be involved in regulating gene expression during the post-meiotic phase of spermatogenesis. Further NMR analysis should give an insight into how this protein binds to and bends DNA. However the role of this modification of DNA structure in the function of Sox-5 will require the isolation and characterization of *in vivo* targets.

ACKNOWLEDGEMENTS

REFERENCES

1. McLaren, A. (1988) *Trends Genet.*, **4**, 153−157.
2. Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P.N. and Lovell-Badge, R. (1991) *Nature*, **351**, 117−121.
3. Jantzen, H.M., Admon, A., Bell, S.P. and Tjian, R. (1990) *Nature*, **344**, 830−836.
4. Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A.M., Lovell-Badge, R. and Goodfellow, P.N. (1990) *Nature*, **346**, 240−244.
5. Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Munsterberg, A., Vivian, N., Goodfellow, P. and Lovell-Badge, R. (1990) *Nature*, **346**, 245−250.
6. Laudet, V., Stehelin, D. and Clevers, H. (1993) *Nucleic Acids Res.*, **21**, 2493−2501.
7. Van de Wetering, M., Oosterwegel, M., Dooijes, D. and Clevers, H. (1991) *EMBO J.*, **10**, 123−132.
8. Travis, A., Amsterdam, A., Belanger, C. and Grosschedl, R. (1991) *Genes Dev.*, **5**, 880−894.
9. Kelly, M., Burke, J., Smith, M., Klar, A. and Beach, D. (1988) *EMBO J.*, **7**, 1537−1547.
10. Staben, C. and Yanofsky, C. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 4917−4921.
11. Griffiths, R. (1991) *Proc. R. Soc. Lond. Biol.*, **244**, 123−128.
12. Denny, P., Swift, S., Brand, N., Dabhade, N., Barton, P. and Ashworth, A. (1992) *Nucleic Acids Res.*, **20**, 2887.
13. Sockanathan, S., Cohen-Tannoudji, M., Collignon, J. and Lovell-Badge, R. (1993) *Genet. Res. Camb.*, **61**, 149.
14. Denny, P., Swift, S., Connor, F. and Ashworth, A. (1992) *EMBO J.*, **11**, 3705−3712.
15. Willison, K., Lewis, V., Zuckerman, K.S., Cordell, J., Dean, C., Miller, K., Lyon, M.F. and Marsh, M. (1989) *Cell*, **57**, 621−632.
16. Smith, D.B. and Johnson, K.S. (1988) *Gene*, **67**, 31−40.
17. Studier, F.W. (1991) *J. Mol. Biol.*, **219**, 37−44.
18. Read, C.M., Cary, P.D., Crane-Robinson, C., Driscoll, P.C. and Norman, D.G. (1993) *Nucleic Acids Res.*, **21**, 3427−3436.
19. Liu, J.H., Gartenberg, M.R. and Crothers, D.M. (1986) *Cell*, **47**, 995−1005.
20. Kim, J., Zwieb, C., Wu, C. and Adhya, S. (1989) *Gene*, **85**, 15−23.
21. Thompson, J.F. and Landy, A. (1988) *Nucleic Acids Res.*, **16**, 9687−9705.
22. Bellve, A.R. (1979) In C.A. Finn (ed.), *Oxford Reviews of Reproductive Biology*. Oxford University Press, Oxford, Vol.1, pp. 159−261.
23. Nasrin, N., Buggs, C., Kong, X.F., Carnazza, J., Goebl, M. and Alexander, B.M. (1991) *Nature*, **354**, 317−320.
24. Harley, V.R., Jackson, D.I., Hextall, P.J., Hawkins, J.R., Berkovitz, G.D., Sockanathan, S., Lovell-Badge, R. and Goodfellow, P.N. (1992) *Science*, **255**, 453−456.
25. Giese, K., Cox, J. and Grosschedl, R. (1992) *Cell*, **69**, 185−195.
26. Harley, V.R., Lovell-Badge, R. and Goodfellow, P.N. (1994) *Nucleic Acids Res.*, **22**, 1500−1501.
27. Giese, K., Pagel, J. and Grosschedl, R. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 3368−3372.
28. Giese, K., Amsterdam, A. and Grosschedl, R. (1991) *Genes Dev.*, **5**, 2567−2578.
29. van de Wetering, M., Oosterwegel, M., van Norren, K. and Clevers, H. (1993) *EMBO J.*, **12**, 3847−3854.
30. Lilley, D.M. (1992) *Nature*, **357**, 282−283.
31. Wu, H.M. and Crothers, D.M. (1984) *Nature*, **308**, 509−513.
32. Kerppola, T.K. and Curran, T. (1991) *Science*, **254**, 1210−1214.
33. Ferrari, S., Harley, V.R., Pontiggia, A., Goodfellow, P.N., Lovell-Badge, R. and Bianchi, M.E. (1992) *EMBO J.*, **11**, 4497−4506.
34. Olins, D.E., Bryan, P.N., Harrington, R.E., Hill, W.E. and Olins, A.L. (1977) *Nucleic Acids Res.*, **4**, 1911−1931.
35. Compton, L.A. and Johnson Jr, W.C. (1986) *Analytical Biochemistry*, **155**, 155−167.
36. van Houte, L., van Oers, A., van de Wetering, M., Dooijes, D., Kaptein, R. and Clevers, H. (1993) *J. Biol. Chem.*, **268**, 18083−18087.
37. Weir, H.M., Kraulis, P.J., Hill, C.S., Raine, A.R., Laue, E.D. and Thomas, J.O. (1993) *EMBO J.*, **12**, 1311−1319.
38. Ner, S.S. (1992) *Curr. Biol.*, **2**, 208−210.