The EMBO Journal vol.13 no.24 pp.6115–6124, 1994

Sex-reversing mutations affect the architecture of SRY–DNA complexes

Andrea Pontiggia[1], Rebecca Rimini[1], Vincent R.Harley[2], Peter N.Goodfellow[2], Robin Lovell-Badge[3] and Marco E.Bianchi[1,4,5]

[1]DIBIT, San Raffaele Scientific Institute, via Olgettina 58, 20132 Milano, Italy, [2]Department of Genetics, Cambridge University, Downing Street, Cambridge CB2 3EH, [3]Laboratory of Eukaryotic Molecular Genetics, MRC National Institute of Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK and [4]Dipartimento di Genetica e di Biologia dei Microrganismi, Università degli Studi, via Celoria 26, 20133 Milano, Italy

[5]Corresponding author

Communicated by G.Magni

The testis determining factor, SRY, is a DNA binding protein that causes a large distortion of its DNA target sites. We have analysed the biochemical properties of the DNA binding domains (HMG-boxes) of mutant SRY proteins from five patients with complete gonadal dysgenesis. The mutant proteins fall into three categories: two bind and bend DNA almost normally, two bind inefficiently but bend DNA normally and one binds DNA with almost normal affinity but produces a different angle. The mutations with moderate effect on complex formation can be transmitted to male progeny, the ones with severe effects on either binding or bending are *de novo*. The angle induced by SRY depends on the exact DNA sequence and thus adds another level of discrimination in target site recognition. These data suggest that the exact spatial arrangement of the nucleoprotein complex organized by SRY is essential for sex determination.

Key words: chromatin/DNA bending/HMG-box/protein–DNA interactions/sex determination

Introduction

In mammals, sex determination requires the Y-chromosome gene *SRY* (Koopman *et al.*, 1991; Sinclair *et al.*, 1990). *SRY* is expressed in the genital ridge of the embryo in a specific time window (between days 10.5 and 12 *post coitum* in the mouse) and induces its differentiation into the testis (Gubbay *et al.*, 1990). The SRY gene product is a DNA binding protein which recognizes sequence-specifically sites that conform to the sequence A/TAACAAA/T (Harley *et al.*, 1992, 1994). The DNA binding activity is essential for testis development (Nasrin *et al.*, 1991; Harley *et al.*, 1992). Thus, SRY protein might exert its effect on development via the transcriptional regulation of the expression of (as yet unidentified) specific genes, possibly by binding to their transcriptional control elements and interacting directly or indirectly with RNA polymerase II. However, this simple picture is inadequate for at least two reasons: SRY proteins have limited selectivity in sequence discrimination and numerous proteins exist that share an extremely similar or identical sequence specificity with SRY.

The DNA binding properties of SRY are specified by its DNA binding domain, which is ~70 amino acids and belongs to the HMG-box class (for a review, see Bianchi, 1995). Its interaction with DNA occurs mainly across the minor groove (van de Wetering and Clevers, 1992; Giese *et al.*, 1994), where there are relatively few hydrogen bond donors and acceptors and the recognition of bases is therefore restricted. This limits the binding specificity of SRY. Numerous substitutions in the target site are allowed (Harley *et al.*, 1992) and specific and non-specific binding affinities differ by a factor of only 20–50 (Ferrari *et al.*, 1992). Therefore, the concentration of SRY necessary to occupy a specific site in the mammalian genome is high, since numerous other sites will compete for SRY binding. This problem is compounded by the presence of SOX proteins which contain HMG-box domains closely related to SRY's and recognize essentially the same set of sequences with very similar affinities (Denny *et al.*, 1992a,b) and therefore compete for occupancy of the same critical site.

Specificity in the control of gene expression can be sharply improved by protein–protein contacts between multiple DNA binding proteins, that cooperatively affect each other's probability of occupying a specific DNA segment. The amino acid sequences of mammalian SRY proteins from even closely related species, however, diverge completely outside the HMG-box (Tucker and Lundrigan, 1993; Whitfield *et al.*, 1993). Moreover, the only mutations in *SRY* that cause sex-reversal in humans always fall within the HMG-box domain (Berta *et al.*, 1990). It is unlikely then for SRY to contain domains (other than the HMG-box itself) devoted to contact other factors.

An alternative possibility exists: SRY might promote contact between other DNA binding proteins by acting via the DNA molecule (see Figure 1) and thus belong to the growing family of factors that modulate transcription by acting as architectural components of stereospecific enhancer complexes (reviewed by Tjian and Maniatis, 1994). Upon binding, human SRY distorts the DNA to dramatic extents: it produces a bend of ~80° centred at the GAACAAAG sequence in the CD3ε enhancer (Ferrari *et al.*, 1992). Proteins bound at either side of the SRY binding site may then be brought into close contact and form a stable complex. In turn, the complex would freeze the deformation of the DNA molecule and indirectly stabilize the binding of SRY. In fact, we have shown previously that SRY will bind with high affinity to distorted DNA molecules, like four-way junctions, irrespective of their sequence (Ferrari *et al.*, 1992).

To find evidence for a possible role of SRY's DNA

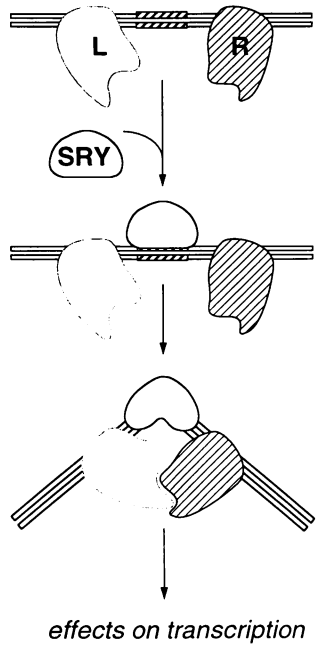

Fig. 1. A model for the action of SRY protein. The human testis determining factor SRY is a DNA binding protein belonging to the HMG-box class. SRY is most closely related to a number of gene-specific transcriptional regulators, which recognize specific sequences in DNA. However, the recognition takes place mainly through the minor groove and has a low selectivity. In addition, SRY proteins even from closely related mammals share similarity only within the HMG-box DNA binding domain and diverge considerably outside of it. This argues against the presence of additional domains involved in transcriptional transactivation. However, SRY bends the DNA considerably after binding to some sites and this may promote protein–protein interactions between factors (L for 'left' and R for 'right') bound to either side of the SRY binding site. Thus, a mechanical action on DNA may induce the formation of a nucleoprotein complex endowed with high specificity (requiring at least three different sites bound by different proteins), high stability (the protein–protein and protein–DNA interactions would be cooperative and reinforce each other) and biological activity (through domains present on the L and R proteins). The experiments reported in this paper support the notion that the deformation induced by SRY must be geometrically precise to serve as a genetic switch.

bending activity in gene expression and differentiation, we have asked whether it was altered in SRY mutants associated with sex-reversal in humans. Five mutations with varying degrees of severity were considered. Mutations F109S and I90M were associated with complete gonadal dysgenesis in the patients where they were first identified, but were also present in normal male relatives of the patients, including the father (Hawkins *et al.*, 1992b; Jäger *et al.*, 1992). These mutations therefore have low penetrance, are transmissible and may cause a differentiation defect only in specific environments or in association with specific genetic backgrounds. Mutations G95R and M64I were identified only in patients with complete gonadal dysgenesis but not in their relatives and have arisen *de novo* (Berta *et al.*, 1990; Hawkins *et al.*, 1992a). Mutation K106I was also associated with complete gonadal dysgenesis, but no male relatives of the patient were available for testing (Hawkins *et al.*, 1992b).

The HMG-boxes from four mutant SRY proteins of sex-reversed patients have variable affinity towards the CD3ε enhancer site, but produce complexes in which the DNA is bent normally. Therefore, the naive expectation that the bending capacity depends on a tight association of the protein to the DNA does not hold. A fifth mutant with mildly reduced DNA binding affinity produces a different bend. This finding is in accordance with the idea that the variation in the geometry of the SRY protein–DNA complex directly causes sex-reversal.

Finally, we have also shown that variations in the sequence of the binding site can cause a different bending of the DNA complexed with wild type and mutant SRY proteins. This implies that different sequences can be discriminated on the basis of the geometry of the complex they support, rather than on the basis of the probability of their occupancy.

Taken together, these results suggest that the correct spatial configuration of the complexes formed by SRY is indeed important for gene expression and subsequent cellular differentiation.

Results

Production of wild type and mutant SRY DNA binding domains

To investigate the relationship between DNA binding, bending and biological role of SRY proteins, we have studied the DNA binding properties of normal and mutant human SRY HMG-boxes with single amino acid substitutions. The HMG-box-coding sequences of SRY genes from a normal male and from five different sex-reversed patients (Table I) were recovered by PCR and cloned in the expression vector pT7-7. Upon induction with IPTG, corresponding polypeptides were expressed at high levels in *Escherichia coli* and were purified to homogeneity.

The affinity of normal and mutant HMG-boxes for linear DNA and four-way junctions

The affinities of the normal and the five mutant HMG-boxes for a DNA fragment containing the CD3ε enhancer site were determined in bandshift assays using a limiting amount of labelled DNA and titrating the polypeptides. Mutant HMG-boxes varied extensively in their DNA binding ability, in accordance with a previous report (Harley *et al.*, 1992). The affinity for DNA of mutant F109S was indistinguishable from that of the wild type, as already noted by Jäger *et al.* (1992). The affinity of mutants I90M and M64I was moderately reduced in comparison with the wild type, whereas the affinity of K106I was reduced by more than two orders of magnitude and G95R did not form any complexes under the conditions we used (Table I). Variations of activity in a bandshift assay can be caused by a variation in the intrinsic affinity of the proteins for DNA, or by the presence of a variable proportion of denatured or inactive protein molecules in the preparation. To rule out the latter alternative, we tested the DNA binding activity of various preparations of the same polypeptide and found them identical.

We then measured the binding activity of the wild type and mutant boxes towards a different DNA target, four-way junctions. The HMG-box domains of HMG1 (an abundant chromatin protein), SRY and UBF (a general transcription factor for RNA polymerase I) all recognize four-way junctions structure-specifically and sequence-

Table I. Summary of *SRY* mutation characteristics and properties of the SRY HMG-boxes

Mutation	Type	K_D linear DNA (M)	Angle (°)	K_D four-way junction DNA (M)
wt		2×10^{-8}	76	10^{-8}
F109S	familial	2.5×10^{-8}	76	10^{-8}
I90M	familial	4×10^{-8}	75	10^{-8}
M64I	*de novo*	6.5×10^{-8}	56	10^{-8}
K106I	relatives unavailable	2×10^{-6}	73	10^{-8}
G95R	*de novo*	$\geq 10^{-5}$		10^{-8}

The mutations in *SRY* we have considered were described by Berta *et al.* (1990), Hawkins *et al.* (1992a,b) and Jäger *et al.* (1992). The dissociation constants for complexes between the various SRY HMG-boxes and DNA were calculated by titrating the SRY HMG-box polypeptides against 0.25 nM DNA and assaying the formation of complexes by electrophoretic mobility shift assays (see Materials and methods). 'Linear' refers to the D probe bearing the CD3ε site, 'four-way junction' refers to junction molecule z, which was chosen because the sequences it contains cannot be recognized sequence-specifically by SRY protein (Ferrari *et al.*, 1992). 'Angle' refers to the angle calculated by the algorithm of Ferrari *et al.* (1992) from the circular permutation assays on the set of probes bearing the CD3ε site. The reported values are the mean of at least two assays.

independently (Bianchi *et al.*, 1989, 1992b; Ferrari *et al.*, 1992; Kuhn *et al.*, 1994). The recognition of this unusual DNA structure is related to the functional properties of this class of proteins (Bianchi *et al.*, 1992a; Lilley, 1992) and depends on the structural integrity and correct folding of the HMG-box domain (Falciola *et al.*, 1994). The affinity of the mutant SRY DNA binding domains towards junction z was indistinguishable from that of the wild type (Table I). Therefore, the various proteins were all equally active towards at least one type of DNA molecule, indicating that the amount of inactive polypeptide chains was irrelevant and that the amino acid substitutions did not cause gross misfoldings of the SRY proteins.

The DNA bending activity of normal and mutant HMG-boxes

We next compared the DNA bending ability of mutant boxes using a circular permutation assay—DNA fragments with a distortion in the middle of the molecule have a different electrophoretic mobility compared with DNA fragments of identical length and sequence with a distortion near one end. Although the relation between electrophoretic mobility and conformation is complex (Levene and Zimm, 1989), the assay allows one to map the locus of protein–DNA interaction and to estimate the amount of distortion introduced in DNA (Wu and Crothers, 1984; Liu-Johnson *et al.*, 1986; Thompson and Landy, 1988; Kerppola and Curran, 1991; Ferrari *et al.*, 1992). Measurements are extremely reproducible, with an accuracy close to 0.5°. Figure 2 shows some typical assays, from which we deduced that the locus of bending falls within the GAACAAA sequence in all cases and we calculated that the wild type SRY HMG-box bends the DNA at this site by 75°, while mutant M64I induces a significantly smaller angle, of ~56°. All other mutants, including K106I, bend the DNA by an angle extremely close to the wild type (Table I). Our results indicate clearly that mutant M64I bends the DNA much less than the wild type, but its binding affinity is reduced only mildly, whereas mutant K106I bends the DNA almost as much as the wild type although its DNA binding affinity is reduced by more than two orders of magnitude. Thus, contrary to a widely held preconception, affinity and DNA bending are not correlated and a weakly binding protein can produce a large DNA distortion.

Wild type SRY and the M64I mutant recognize the same repertoire of sequences in linear DNA

The difference in the angle induced by mutant M64I, viewed against the relative uniformity of the other mutants, appears quite large. In comparison, the DNA bending properties of SRY proteins from related animals (gorilla, gibbon, baboon, chimpanzee, pygmy chimpanzee, orang-utan and marmoset) are strictly conserved. We have verified that the deformations they introduce into the CD3ε probe fall within a 8° range, despite several variations in their amino acid sequences (Pontiggia *et al.*, 1995). Thus, it is tempting to consider the difference in bending ability of the M64I mutant as a key to the biological consequences of the mutation. However, a more trivial explanation is possible: mutation M64I might modify the sequence specificity of SRY protein. In order to test this hypothesis, we performed several cycles of selection and amplification of double-stranded DNA from a population of molecules initially displaying all possible sequence variations at the central 10 nucleotide residues. After five cycles of selection/amplification, the M64I mutant HMG-box domain had selected a population of molecules that bound significantly better than the initially degenerate population; additional cycles of selection/amplification did not increase the average affinity of the population for the protein. The selected molecules were cloned, picked at random and sequenced (Figure 3). These sequences can be aligned, are rich in As and Ts and the distribution of bases at seven individual positions is clearly non-random. The consensus that can be derived exactly matches the consensus determined for molecules selected in selection/amplification protocol by wild type SRY protein, A/TAACAAA/T (Harley *et al.*, 1994).

We also compared the binding preferences of both mutant and wild type SRY HMG-boxes with respect to entire repertoires of sequences, rather than individual sequences. A competition between the initial (non-specific) and final (specific) population of DNA molecules was performed (Figure 3). The wild type and M64I proteins were titrated to shift a similar proportion of molecules from the selected population. Addition of similar amounts of the initial random population reduced the bandshifts caused by the wild type and the mutant by comparable amounts. Similar results were also obtained using defined sequences rather than populations. We conclude that the mutant and the wild type

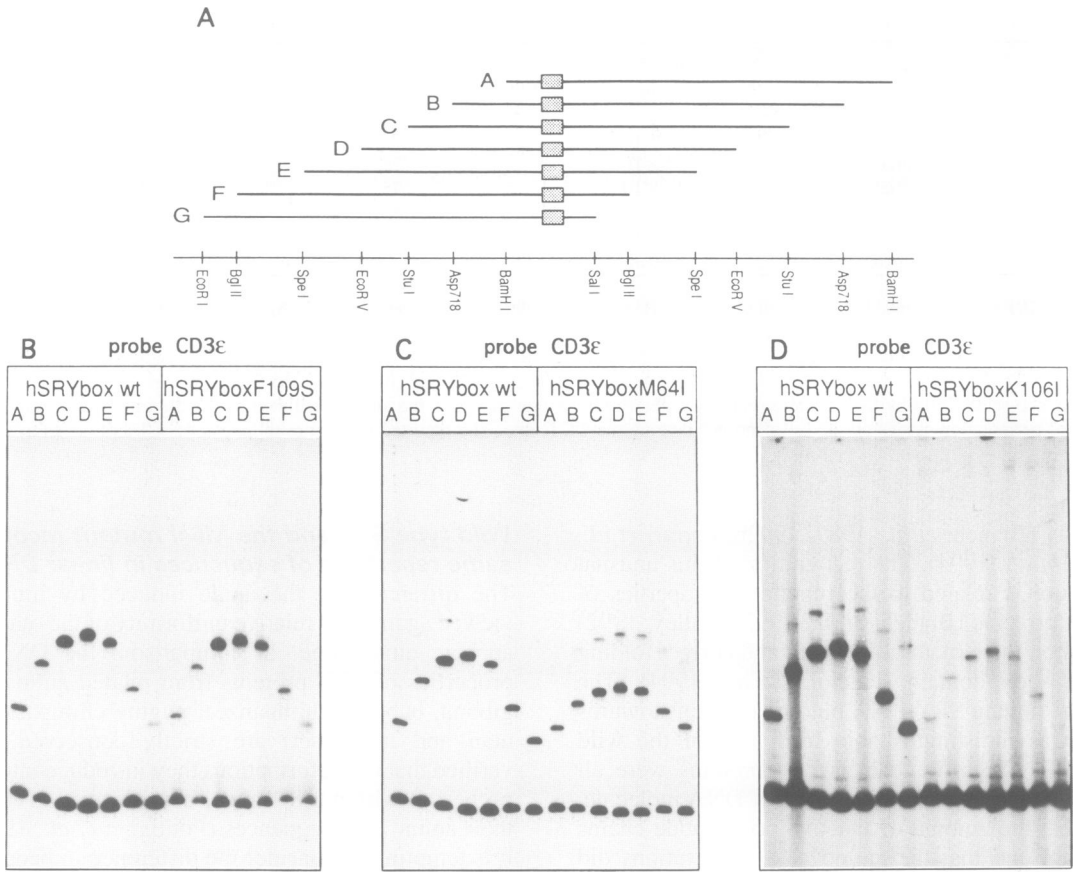

Fig. 2. DNA bending activity of mutant SRY HMG-boxes on the CD3ε sequence. (**A**) DNA probes used for the circular permutation analysis. Plasmid pBend2CD3ε, containing the CD3ε site (hatched box) flanked by tandemly repeated DNA sequences, was cleaved at the restriction sites indicated in the map. The DNA fragments obtained in this way (designated A–G) all contain circular permutations of the same sequence of 141 bp. (B, C and D) Circular permutation analysis of DNA bending induced by mutant hSRYboxes F109S, M64I and K106I. Each panel shows a comparison between the electrophoretic mobilities of complexes of wild type hSRYbox and mutant F109S (**B**), mutant M64I (**C**) and mutant K106I (**D**) with probes A–G containing the CD3ε site GAACAAAG. Three fmol of labelled DNA probes were mixed in 9 μl of standard DNA binding buffer (see Materials and methods). One μl of purified protein (≈2 ng of hSRYbox wild type, ≈2 ng of hSRYbox F109S, ≈10 ng of hSRYbox M64I, ≈25 ng of hSRYbox K106I) was added to the various mixtures. Electrophoresis and autoradiography were performed as indicated in Materials and methods, however, the autoradiography in (D) was overexposed to detect the faint bands of DNA complexed to hSRYbox K106I. The retarded bands running behind the principal protein–DNA complex band are complexes containing more than one polypeptide molecule per DNA molecule. No significant difference in the mobility of the free DNA probes can be observed, indicating that the CD3ε site does not distort DNA on its own.

SRY HMG-boxes recognize the same sites and that each specific 'preferred' sequence is recognized with similar specificity over 'non-preferred' sequences.

DNA sequence affects DNA bending by SRY

The experiments reported in the previous sections establish that one specific sequence is moulded into different configurations by mutant M64I and by the wild type SRY. However, there is no guarantee that the CD3ε site is a true SRY binding site; in fact, a physiological target for SRY has not yet been identified. Moreover, the DNA itself is possibly the major determinant in the final conformation of the nucleoprotein complex and slight variations of the DNA sequence interacting with SRY might affect the geometry of the nucleoprotein complex quite substantially. We therefore measured, by means of the circular permutation assay, the angle of deflection produced by wild type and mutant SRY proteins on two sequences that diverge slightly from the CD3ε sequence.

The sequence TAACAATG has been reported to bind SRY protein with higher affinity than the GAACAAAG sequence present in the enhancer of the CD3ε gene, while the sequence GAACACAG was reported to be a poorer binding site (Giese *et al.*, 1992). The two sequences were substituted for the CD3e sequence in plasmid pBendCD3ε and the circularly permuted fragments obtained from the two new plasmids (called pBend2mut11 and pBend2mut0, respectively) were tested electrophoretically with wild type and mutant SRY HMG-boxes for binding affinity and bending features (Table II).

The fragments bearing the sequence TAACAATG bound every SRY HMG-box variant with ~2-fold better affinity than the GAACAAAG sequence. The circular permutation assay showed that the locus of flexure was the same as for pBendCD3ε fragments, but the bending was slightly less.

The fragments bearing the sequence GAACACAG bound every SRY HMG-box variant ~6-fold less efficiently than the pBendCD3ε fragments. Again, the protein-induced bending was centred around the same locus,

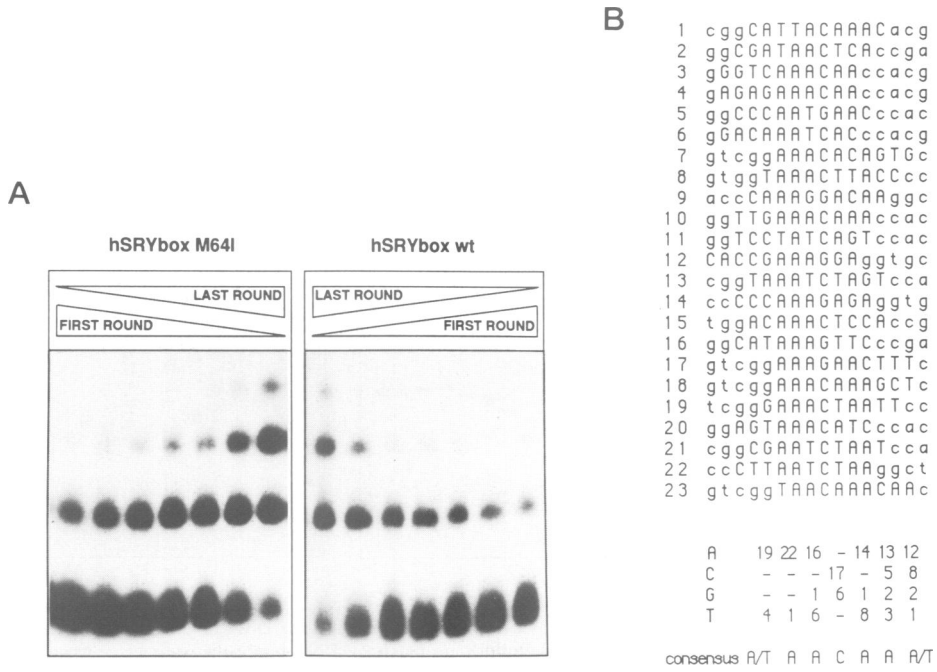

Fig. 3. The HMG-boxes of wt SRY and mutant M64I recognize the same repertoire of sites. (**A**) A population of 60 bp DNA molecules initially containing all possible 10 bp sequences at their centre was subjected to five rounds of selection/amplification by SRY HMG-box M64I (see Materials and methods). At the end of the procedure, labelled templates amplified after the first round of selection by hSRYbox M64I (first round) and after the fifth round (last round) were mixed in different ratios, mixed with wt SRY HMG-box or M64I SRY HMG-box and electrophoresed. The same total amount of DNA (21 fmol) was used as probe in every lane (last round/first round: 0/21, 3/18, 7/14, 10.5/10.5, 14/7, 18/3 and 21/0). To 9 µl of the reaction mix containing the DNA probes, 20 ng of hSRYbox M64I (right panel) or 5 ng of hSRYbox wt (left panel) were added. After incubation on ice, samples were applied to vertical 6.5% polyacrylamide gels in 0.5× TBE as described in Materials and methods. As expected, the population of DNA molecules amplified after the fifth round of selection/amplification with the mutant SRY HMG-box M64I is recognized much better by mutant M64I than the initial population. However, the same is true if the populations are challenged with the wt SRY HMG-box. Mixing the initial and final populations in different proportions is equivalent to performing a competition experiment between a specific binding site and aspecific DNA. As the figure shows, again no difference is apparent between the bandshifts produced by the wt SRY HMG-box and mutant M64I, an indication that their binding selectivity is the same. (**B**) Sequence of some DNA molecules selected after the fifth selection/amplification cycle with SRY HMG-box M64I. The lower-case bases belong to the polylinker of plasmid Bluescript. The frequency of each base at each position is indicated below the sample set of sequences. The derived consensus matches the consensus for wild type SRY (Harley *et al.*, 1994).

showing that the GAACACAG sequence was preferred to any other sequence in the probe. However, the bending was dramatically reduced with the wild type SRY HMG-box, while mutant M64I appeared unable to induce any measurable bending of these fragments (Figure 4).

Discussion

How does SRY work?

The product of the sex-determining SRY gene is a DNA binding protein that controls the differentiation of pre-Sertoli cells during embryonic life. The only conserved part of the SRY protein in related animals is the DNA binding domain, which recognizes a defined repertoire of sites. However, the sequence selectivity of SRY is rather low and indistinguishable from that of other proteins with similar DNA binding domains but very different biological functions. Both we and others have shown that SRY induces a large alteration of the local geometry of the DNA sites to which it binds and that DNA molecules that already contain sharp bends or kinks are actually bound by SRY with high affinity and selectivity, irrespective of the sequences that they contain (Ferrari *et al.*, 1992; Giese *et al.*, 1992). Such unusual properties suggest that the DNA bending function might be central to the biological activity of SRY, as well as of other proteins with similar HMG-box DNA binding domains (Bianchi, 1994; Grosschedl *et al.*, 1994). As illustrated in Figure 1, SRY would promote the interaction of flanking transcription factors by acting mechanically on the DNA molecule. Two predictions are central to the model. (i) SRY does not promote transcription on its own, but depends on the presence of 'left' and 'right' transcription factors, which are only active if physically associated. In fact, known SRYs lack recognizable transcription activation domains and their sequences are completely divergent outside the HMG-box. The existence of the predicted 'left' and 'right' factors rests on the analogy with the LEF-1/TCRα system (Ho and Leiden 1990; Giese *et al.*, 1992; Giese and Grosschedl, 1993). (ii) The left and right transcription factors interact only in the presence of SRY, but not of any other protein. This is possible only if SRY organizes the nucleoprotein complex in a specifically defined way, which cannot be attained spontaneously nor by proteins that can bind the same sites as SRY.

The data we report test the prediction that the geometry of the SRY–DNA complex must be a prime determinant in sex determination. A mutation that alters the geometry of the nucleoprotein complex is associated with a loss of function of SRY. Moreover, different binding sites can be discriminated on the basis of the geometry they adopt when bound by SRY.

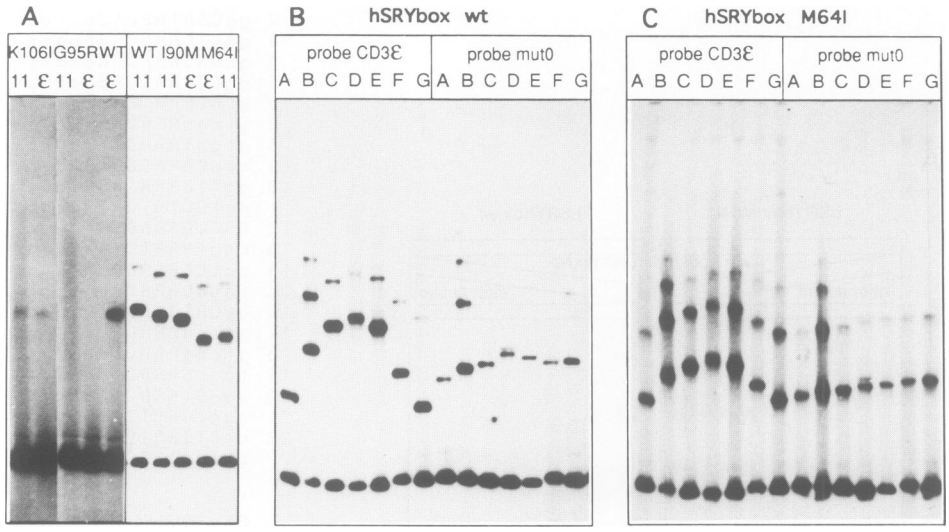

Fig. 4. DNA binding sites with different sequences are bent differently by SRY HMG-boxes. (**A**) Mobility shift assay of wild type and mutant SRY HMG-boxes on CD3ε (ε) and mut11 (11) binding sites. Probe D was used because it contains the DNA binding site (CD3ε or mut11) in the middle of the fragment (Figure 2A) and the differences in electrophoretic mobilities are maximal. DNA binding mixtures (see Materials and methods) were set up to contain 3 fmol of CD3ε D probe or mut11 D probe in 9 μl. To each mixture containing the CD3ε D probe, 1 μl of purified protein was added, containing ≈50 ng of hSRYbox of G95R, ≈3 ng of hSRYbox wild type, ≈10 ng of hSRYbox I90M, ≈15 ng of hSRYbox M64I, ≈25 ng of hSRYbox K106I, respectively. About a quarter as much wild type hSRYbox and half as much for each mutant SRY HMG-box were added to mixtures containing the mut11 D probe. Autoradiographic signals were also analysed with a PhosphorImager. No binding activity was found for hSRYbox G95R. The same results were also obtained when using larger amounts of purified hSRYbox G95R protein. (**B** and **C**) Circular permutation analysis of wild type hSRYbox (B) and mutant M64I (C) on CD3ε and mut0 binding sites. DNA probes A–G (3 fmol) were mixed in 9 μl of standard DNA binding buffer (see Materials and methods). To the various mixtures containing labelled DNA fragments 1 μl of purified protein solution was added, containing ≈2 ng of hSRYbox wild type (probe CD3ε), ≈6 ng of hSRYbox wild type (probe mut0), ≈15 ng of hSRYbox M64I (probe CD3ε), ≈60 ng of hSRYbox M64I (probe mut0). The samples were electrophoresed and autoradiographed as described in Materials and methods. The retarded bands running behind the principal protein–DNA complex band are complexes containing more than one polypeptide molecule per DNA molecule. Unbound DNA probes containing the mut0 site have the same electrophoretic mobility, indicating that the mut0 site does not distort DNA on its own.

Table II. Influence of the DNA sequence on the amplitude of the angle induced by SRY proteins.

		wt		M64I	
Probe	Sequence	K_D (M)	Angle (°)	K_D (M)	Angle (°)
CD3ε	GAACAAAG	2×10^{-8}	76	6.5×10^{-8}	56
mut11	TAACAATG	10^{-8}	73	3×10^{-8}	55
mut0	GAACACAG	10^{-7}	35	3.5×10^{-7}	<10

The dissociation constants for complexes between SRY HMG-boxes and DNA were calculated by titrating the SRY HMG-box polypeptides against 0.25 nM DNA and assaying the formation of complexes by electrophoretic mobility shift assays (see Materials and methods). 'Angle' refers to the angle calculated by the algorithm of Ferrari *et al.* (1992) from the circular permutation assays on the set of probes bearing the CD3ε site. The reported values are the mean of at least two assays.

Dissection of the DNA binding and DNA bending activities of SRY protein

The most obvious result of our analysis is that the DNA binding and DNA bending activities of the SRY protein can be separated. The K106I amino acid substitution in SRY reduces the DNA binding affinity of the protein >100-fold and yet still allows the bending of the DNA target to the same extent attained by the wild type. Conversely, the M64I amino acid substitution reduces the DNA binding affinity only 3-fold, but causes a reduction in DNA bending from ~75° to ~56°.

The non-correlation between the binding affinity of the SRY variants and the extent of deformation caused in DNA is at first sight surprising. How can a weakly binding protein like SRY K106I cause a large deformation in DNA? The simple answer is that mutant K106I binds weakly precisely because it still significantly deforms DNA. The binding affinity of a protein to its ligand is related to the difference in free energy at equilibrium between the bound and unbound states of the system that comprises the protein, the ligand and the aqueous environment. The free energy variation between the bound and unbound states of the SRY/DNA system can be decomposed (somewhat arbitrarily) in several contributions, as shown in the diagram in Figure 5. One contribution corresponds to the establishment of chemical bonds between SRY and the nucleic acid, irrespective of the sequence or the conformation of the DNA, in a non-specific mode. A second contribution corresponds to the formation of sequence-specific contacts between SRY and a favoured binding site. The third contribution, the deformation of the DNA site, is negative: DNA bending

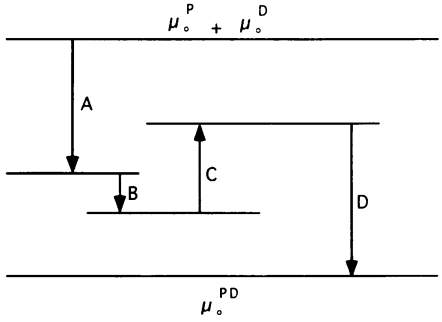

Fig. 5. Thermodynamics of the binding of SRY HMG-boxes to DNA. The chemical potential of SRY protein and of a linear DNA molecule containing a preferred target site are indicated as μ_o^P and μ_o^D, respectively; the chemical potential of the complex as μ_o^{DP}. The difference in the energy levels $\mu_o^P+\mu_o^D$ and μ_o^{DP} represents the $\Delta G_o'$ variation in free energy associated with complex formation. The total free energy variation can be divided in several contributions: (**A**) the free energy variation associated with non-specific protein–DNA interactions; (**B**) that associated with sequence-specific interactions; (**C**) the one associated with the deformation of the DNA; (**D**) the one associated with conformation-specific interactions. The formation of the complex is thermodynamically favoured so long as A + B + D is larger in absolute value than C.

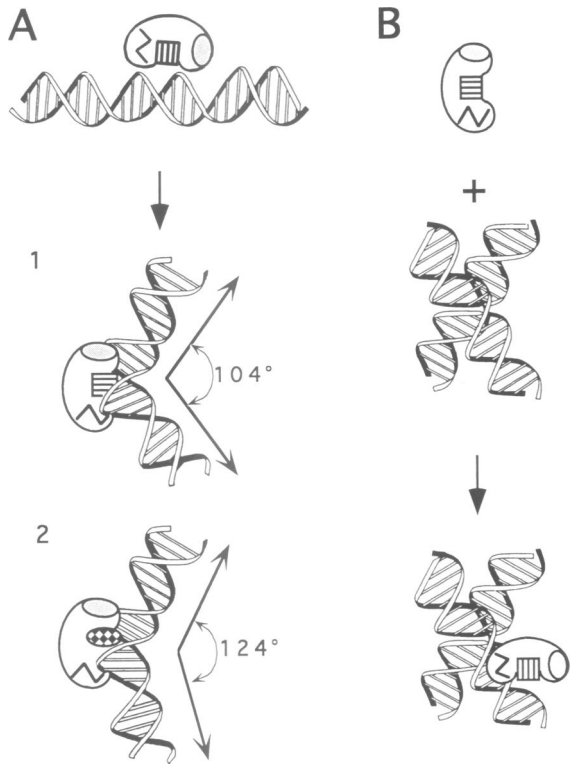

Fig. 6. A schematic model for the interaction of SRY proteins with linear and four-way junction DNAs. The DNA binding domain of wild type SRY is depicted here as a rounded object with three sites which are important for the interaction with DNA. The grey disc is involved in sequence-specific recognition, the zig-zag is associated with conformation-specific interactions and the square with bars represents the site in the protein responsible for the deflection of linear DNA. In reality, each site may be composed of multiple residues; conversely, some residues may contribute to more than one of the sites that are shown here as independent. (**A**) Interaction with linear DNA. When wild type SRY interacts with a specific sequence in linear DNA (1), all three sites will establish contacts with the DNA in the complex and the DNA deflection will be mainly determined by the burrowing of the site represented as the square with bars into the minor groove. The substitution of a residue in this site (chequered disc) will modify marginally the overall affinity of the protein for the DNA, but will have a profound effect on the overall geometry of the complex (2). (**B**) Interaction with four-way junctions. The affinity of SRY for four-way junctions does not depend on the sequence of the DNA. Moreover, the DNA already contains kinks. Therefore, we expect that the sequence-specific site and the DNA-deflecting site will not play important roles in this type of interaction, which will be mainly governed by the conformation-specific site. To stress this point, in the drawing the grey disc and the square with bars are drawn as not making contacts with the four-way junction, although of course this may not reflect the real situation.

takes up energy. A last contribution represents shape-specific interactions between the protein and the DNA, once this is deformed. If the mutant K106I deforms the DNA in the same way as the wild type, the negative contribution will be the same. Then, if the mutation reduces any one of the other contributions, the sum of the free energy variations of all the contributions will be close to zero and the protein will bind weakly to its ligand.

A second key result of our analysis is that the affinity of the various mutated SRY HMG-box domains for specific sites on linear DNA is variable, but the affinity for the four-way junction DNA is identical. This implies that the mechanistic details of duplex binding and junction binding are different, at least partially. We suggest that the SRY mutations we have analysed do not affect the shape-recognizing properties of the HMG-box, but affect one of the steps by which the linear DNA molecule is moulded into a specific shape. On the biological level, the equal recognition of four-way junctions by wild type and mutant SRY proteins formally rules out distorted structures as the physiological target of SRY and suggests that sequence recognition is essential although probably not sufficient for the action of SRY in differentiation.

The model in Figure 6 summarizes our current thinking on SRY binding to DNA. We suggest that the DNA binding domain of SRY contains sites that allow sequence discrimination in linear DNA, sites that deflect the DNA and sites that establish conformation-specific contacts with bent DNA. The correspondence between sites and specific amino acid residues need not be univocal: each site will be made up of several residues forming a surface and a certain residue might contribute to more than one site at the same time. All three sites must be involved in the establishment of protein–DNA contacts in complexes containing the SRY HMG-box and the distorted double helix; mutations affecting any of these sites have the potential to alter the affinity of the protein for a specific target sequence in linear DNA. However, only amino acid substitutions in the DNA deflecting site will modify the geometry of the nucleoprotein complex. Thus, M64 must be one of the residues involved in DNA distortion. The interaction with four-way junctions must mimic at least partly the interaction with double-stranded DNA and probably involves the same conformation-specific contacts. The site for sequence-specific recognition cannot contribute much to the interaction with four-way junctions, since this is largely sequence insensitive (Ferrari *et al.*, 1992). Likewise, the site responsible for DNA distortion might not play an important role in the interaction with

the four-way junction, which is already distorted and is predicted to have a widened minor groove (Lilley and Clegg, 1993). The fact that none of the amino acid substitutions we examined alters the affinity of SRY for four-way junctions indicates that none of the affected residues belong to the conformation-specific site.

The contribution of the DNA sequence to the organization of SRY-induced nucleoprotein complexes

As discussed in the previous paragraph, the biological activity of SRY appears to depend critically on the recognition of specific sequences. We have shown that target sites with altered sequences can still be recognized by SRY, but will be bent differently. This offers a solution to the puzzle of the limited sequence specificity of SRY, as opposed to the required sequence-specificity of its action. A certain DNA sequence may turn out to be a key element in the regulation of gene expression not because SRY binds especially well to it, but because the SRY-induced angle is the correct one for the establishment of a productive complex involving other proteins (see Figure 1). In fact, different HMG-box proteins may recognize the same sequence, but they will produce different angles: while human SRY induces an angle of ~80°, LEF-1 produces an angle of ~130° (Giese *et al.*, 1992).

At a mechanistic level, the intrinsic bendability of specific sequences may ultimately dictate whether SRY will bind to them or not, as suggested by King and Weiss (1993) and how much the DNA will be bent in the complex. Studies on the adaptation of double-helical DNA to the curved path around nucleosomes have indicated that different sequences indeed have different flexibility (reviewed by Travers, 1988). In any event, it is clear that the issues of bendability and angle-choice taken together amount to the likely existence of yet another informational level in DNA readout, one that ultimately depends on the sequence but in quite indirect and complex ways.

Correlation between the DNA binding properties of the SRY HMG-box mutants and the phenotype of the affected patients

In this study, we have adopted the classical genetic approach of understanding a phenomenon by considering mutations that perturb it. To what extent, however, is there a causal relationship between the chemico-physical properties of the DNA binding domains of mutant SRY proteins and the lack of testes formation? A DNA binding protein must bind to DNA if it is to work and in this respect the biological consequences of mutations G95R and K106I correspond well to the reduced binding affinity of correponding protein, as had already been noted in a previous study (Harley *et al.*, 1992). At the other extreme, mutations F109S and I90M can be transmitted from father to son and are indeed present in a number of physiologically differentiated males: they clearly behave as conditional mutations. From this point of view, it is quite logical that the mutated proteins differ little from wild type SRY with respect to their DNA binding and DNA bending properties. Their failure to function in some cases, but not all, must depend on the genetic background or the environment. It has already been shown that the genetic background is important in sex determination: in the mouse *Mus musculus,* when the Y chromosome of the variant *poschiavinus* is present on a C57BL6 background, XY females result, but the same chromosome gives males in all other backgrounds.

Mutant M64I is at the same time more informative and more difficult to interpret. This mutation is *de novo*, which suggests that it is always associated with sex-reversal. The 3-fold reduction in DNA binding affinity with respect to the wild type is unimpressive and it is unlikely that a reduction of this extent might completely subvert the controlled physiological phenomenon. On the other hand, a reduction in the deformation of the DNA translates into a significant dislocation of proteins bound a few helical turns from the SRY binding site (see Figures 1 and 6). We cannot measure the exact difference in the architecture of the SRY–DNA complex responsible for sex determination for two reasons: circular permutation is a low-resolution assay and no bona fide target site for SRY has yet been identified. The first limitation is not very relevant, since in any event the differences in DNA distortion are large enough to be identified qualitatively. The second limitation is more critical, but we point out that M64I always distorts the DNA differently from the wild type, whatever the sequence of the target site.

There is another limit in our understanding of the role of DNA distortion in SRY action: we do not know how tolerant the system is to perturbation. To obtain a rough idea of the degree of wobble allowed in SRY-induced nucleoprotein complexes, we compared the deflection angles induced by human SRY and its primate and mouse counterparts. Primate SRY proteins (gorilla, gibbon, baboon, chimpanzee, pygmy chimpanzee, orang-utan and marmoset) all bend the AACAAAG site in the CD3ε enhancer in a very similar way to human SRY (the maximum difference is 6°), despite the accumulation of several amino acid substitutions in the DNA binding domain (Pontiggia *et al.*, 1995). Such evolutionary conservation suggests a very limited permissible wobble. On the other hand, the angle produced by mouse SRY is quite different from that produced by human SRY (Giese *et al.*, 1994) and this correlates with the inability of transgenic human *SRY* to cause male differentiation in mice (Koopman *et al.*, 1991). The angle variation may be the key molecular defect of mutant M64I, although this deduction is not supported by direct experiments on transcription control. However, we have excluded the possibility that the M64I SRY mutant causes gonadal dysgenesis because its sequence specificity is different from that of the wild type: both proteins select a comparable repertoire of binding sequences in selection/amplification experiments.

In conclusion, we have shown that the DNA binding and DNA bending activities of SRY protein can be dissociated by mutation. The DNA bending activity of SRY is also strongly affected by the underlying sequence of the binding site. We have also shown that the defect of one mutant SRY protein might be traced to its anomalous DNA bending activity. Our findings are in keeping with the hypothesis of a direct role of protein-induced DNA bending in certain types of transcriptional control and should stimulate increased efforts in the study of DNA

structural plasticity and in the search for proteins that modulate such plasticity.

Materials and methods

Construction of the plasmid series pT7-SRYboxes

Oligonucleotides SRYboxdir (CCACATATGCAGGATAGAGTGAAGCGA) and SRYboxrev (CGAAGCTTAACGACGAGGTCGATACTT) were synthesized by the phosphotriester method and were used for PCR without purification. PCR mixtures (50 µl) contained 50 pmol each of oligonucleotides SRYboxdir and SRYboxrev, 0.2 mM dNTPs, 400 ng purified human total genomic DNA, 1 U *Taq* DNA polymerase and 5 µl *Taq* polymerase 10× buffer (Perkin-Elmer Cetus). Thirty cycles of denaturation (30 s at 94°C), annealing (60 s at 50°C) and polymerization (90 s at 72°C) were performed on a Perkin-Elmer Cetus instrument. The reaction products were cleaved with restriction endonucleases *Nde*I and *Hin*dIII, gel purified, ligated between the *Nde*I and *Hin*dIII sites of plasmid PT7-7 (Tabor and Richardson, 1985) and cloned into *Escherichia coli* DH5α. The resulting plasmids, pT7-SRYbox-wt, -M64I, -I90M, -G95R, -K106I and -F109S, were checked by sequencing with T7 DNA polymerase and were then introduced into strain BL21(DE3) (Studier *et al.*, 1991).

Purification of SRY HMG-boxes

The different SRY HMG-boxes were synthesized in BL21(DE3) cells harbouring the pT7-SRYbox plasmids. Cells were grown, induced, harvested and sonicated as described for the preparation of peptide HMG1bA (Bianchi *et al.*, 1992b). Nucleic acids were removed by batch absorption to DEAE-cellulose at 0.45 M NaCl. Each extract was centrifuged at 10 000 r.p.m. in a SS34 rotor for 20 min at 4°C. The supernatant was filtered through a cache of glass wool, applied to a MonoS HR5/5 FPLC column (Pharmacia) equilibrated with buffer D (HEPES pH 7.9, 0.2 mM EDTA). The column was eluted with a linear gradient from 0 to 2 M NaCl in buffer D. SRY HMG-boxes were eluted in fractions containing ~1 M NaCl. The proteins appeared to be >99% pure. Titrations of pure SRY HMG-boxes were first performed on polyacrylamide−SDS−tricine gels by Coomassie blue staining; exact concentrations were then determined by measuring the A_{280} of serial dilutions of pure proteins. Fractions were kept frozen at −20°C without apparent loss of activity.

Electrophoretic mobility shift assays

DNA binding buffer (10 µl of final volume) contained 8% Ficoll, 100 mM NaCl, 10 mM HEPES pH 7.9. To these components we added, in various combinations, synthetic four-way junction DNAs or linear duplex DNAs (final concentrations are indicated in the legends to the figures). Purified polypeptides were incorporated in the indicated amounts. The usual order of addition was (i) fixed components, (ii) proteins, (iii) DNAs. After incubation for 10 min on ice, samples (5 µl) were applied to vertical 10% polyacrylamide gels in 0.5× TBE and electrophoresed at 11 V/cm. The gel was dried and autoradiographed with Kodak XAR-5 films at −80°C for 24–72 h.

Determination of complex dissociation constants

Wild type and mutant SRY HMG-boxes were titrated into binding mixtures containing a fixed amount (0.25 nM) of labelled cruciform z or labelled fragment D from pBend2CD3ε, pBend2mut0 or pBend2mut11. Samples were electrophoresed at 4°C in 10% polyacrylamide gels as described in the previous paragraph. The radioactivity present in the bands was measured by exposing the wet gel to PhosphorImager screens (Molecular Dynamics). Under conditions of protein excess, the dissociation constant is equivalent to the concentration of polypeptide where half of the input DNA is taken up in the complexes and half is free.

Circular permutation assays

Plasmids pBend2 containing DNA sequences CD3ε, mut11 and mut0 were prepared by insertion of annealed synthetic oligonucleotides (CTAGAGAGCGCTTTGTTCTCAG and TCGACTGAGAACAAAGCGCTCT for pBend2CD3ε, CTAGAGAGCGCATTGTTATCAG and TCGACTGATAACAATGCGCTCT for pBend2mut11, CTAGAGAGCGCTGTGTTCTCAG and TCGACTGAGAACACAGCGCTCT for pBend2mut0) between the *Xba*I and *Sal*I restriction sites in plasmid pBend2. Preparation of probes, electrophoresis and data analysis were carried out as previously reported (Ferrari *et al.*, 1992).

DNA binding site selection

Sequences binding to wt and M64I SRY HMG-boxes were selected from a pool of double-stranded 60 bp DNA molecules containing a central stretch of 10 random base pairs, flanked by two unique sequences of 25 bp each. The sequences of the oligonucleotides used were as follows: random 60mer template, CTGGTCGGGTGAATTCGTGTCGTGGN$_{10}$CCGACCCAGCGAATTCAGAGCATG; upper primer, CTGGTCGGGTGAATTCGTGTCGTGG; lower primer ACATGCTCTGAATTCGCTGGGTCGG. The first step of selection was performed using a ^{32}P-labelled pool of double-stranded molecules (≈10^8 c.p.m./µg) prepared by extending the lower primer annealed to the random 60mer template. One µg of lower primer was mixed with 300 ng of 60mer random template in 14 µl of annealing buffer (T7 sequencing kit from Pharmacia) and annealed according to the manufacturer's instructions. T7 DNA polymerase (8 U), 3 µl of labelling mix and 30 µl of [α-^{32}P]dCTP were added and incubated for 30 min at 37°C. Fifty ng of the population of double-stranded DNA molecules were incubated with different amounts of protein for 10 min on ice. Samples (5 µl) were applied to vertical 7% polyacrylamide gels in 0.5× TBE and electrophoresed at 11 V/cm at room temperature. The gel was autoradiographed with Kodak XAR-5 film at 4°C for a few hours. The lanes where the shifted band contained <2% of the total DNA were identified: the shifted band was recovered by cutting out a gel slice, crushing it and eluting the DNA in 0.1 ml of water. The eluted DNA was amplified by PCR using 'upper' and 'lower' primers labelled with T4 kinase and [γ-^{32}P]ATP. The PCR mixture (50 µl) contained 50 pmol of each primer, 0.2 mM dNTPs, 10 µl of eluted DNA solution, 1 U *Taq* DNA polymerase and 5 µl *Taq* polymerase 10× buffer (Perkin-Elmer Cetus). Thirty five cycles of denaturation (60 s at 94°C), annealing (60 s at 55°C) and polymerization (30 s at 72°C) were performed on a Perkin-Elmer Cetus instrument. The amplified DNA was gel-purified, incubated with proteins and electrophoresed as described. Five rounds of selection/amplification were sufficient to select a population with maximal affinity for the protein. The products of the fifth round of selection were subcloned into the *Eco*RV site of pBluescript (Stratagene) and sequenced.

Acknowledgements

We are grateful to S.Ottolenghi for critically reading the manuscript and to all our colleagues for helpful discussions. R.R. is supported by a fellowship from Fondazione San Romanello del Monte Tabor, V.H. and P.N.G. are supported by the Wellcome Trust. M.E.B. is a recipient of Human Capital Mobility grant ERBCHRXCT940482.

References

Berta,P., Hawkins,J.R., Sinclair,A.H., Taylor,A., Griffiths,B.L., Goodfellow,P.N. and Fellous,M. (1990) *Nature*, **348**, 448–450.

Bianchi,M.E. (1995). In Lilley,D.M.J. (ed.), *DNA–Protein Structural Interactions*. Oxford University Press, Oxford, in press.

Bianchi,M.E., Beltrame,M. and Paonessa,G. (1989) *Science*, **243**, 1056–1059.

Bianchi,M.E., Beltrame,M. and Falciola,L. (1992a) In Eckstein,F. and Lilley,D.M.J. (eds), *Nucleic Acids and Molecular Biology*. Springer Verlag, Berlin, Vol. 6, pp. 112–128.

Bianchi,M.E., Falciola,L., Ferrari,S. and Lilley,D.M.J. (1992b) *EMBO J.*, **11**, 1055–1063.

Denny,P., Swift,S., Brand,N., Dabhade,N., Barton,P. and Ashworth,A. (1992a) *Nucleic Acids Res.*, **20**, 2887.

Denny,P., Swift,S., Connor,F. and Ashworth,A. (1992b) *EMBO J.*, **11**, 3705–3712.

Falciola,L., Murchie,A.I.H., Lilley,D.M.J. and Bianchi,M.E. (1994) *Nucleic Acids Res.*, **22**, 285–292.

Ferrari,S., Harley,V.R., Pontiggia,A., Goodfellow,P.N., Lovell-Badge,R. and Bianchi,M.E. (1992) *EMBO J.*, **11**, 4497–4506.

Giese,K., Cox,J. and Grosschedl,R. (1992) *Cell*, **69**, 185–195.

Giese,K. and Grosschedl,R. (1993) *EMBO J.*, **12**, 4667–4676.

Giese,K., Pagel,J. and Grosschedl,R. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 3368–3372.

Grosschedl,R., Giese,K. and Pagel,J. (1994) *Trends Genet.*, **10**, 94–100.

Gubbay,J., Collignon,J., Koopman,P., Capel,B., Economou,A., Münsterberg,A., Vivian,N., Goodfellow,P.N. and Lovell-Badge,R. (1990) *Nature*, **346**, 245–250.

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.

Harley,V.R., Lovell-Badge,R. and Goodfellow,P.N. (1994) *Nucleic Acids Res.*, **22**, 1500–1501.
Hawkins,J.R., Taylor,A., Berta,P., Levilliers,J., Van der Auvera,B. and Goodfellow,P.N. (1992a) *Hum. Genet.*, **88**, 471–474.
Hawkins,J.R., Taylor,A., Goodfellow,P.N., Migeon,C. J., Smith,K.D. and Berkovitz,G.D. (1992b) *Am. J. Hum. Genet.*, **51**, 979–984.
Ho,I.C. and Leiden,J.M. (1990) *Mol. Cell. Biol.*, **10**, 4720–4727.
Jäger,R.L., Harley,V.R., Pfeiffer,R.A., Goodfellow,P.N. and Scherer,G. (1992) *Hum. Genet.*, **90**, 350–355.
Kerppola,T.K. and Curran,T. (1991) *Science*, **254**, 1210–1214.
King,C.-Y. and Weiss,M.A. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 11990–11994.
Koopman,P., Gubbay,J., Vivian,N., Goodfellow,P. and Lovell-Badge,R. (1991) *Nature*, **351**, 117–121.
Kuhn,A., Voit,R., Stefanovsky,V., Evers,R., Bianchi,M.E. and Grummt,I. (1994) *EMBO J.*, **13**, 416–424.
Levene,S.D. and Zimm,B.H. (1989) *Science*, **245**, 396–399.
Lilley,D.M.J. (1992) *Nature*, **357**, 282–283.
Lilley,D.M.J. and Clegg, R.M. (1993) *Annu. Rev. Biophys. Biomol. Struct.*, **22**, 299–328.
Liu-Johnson,H.-N., Gartenberg,M.R. and Crothers,D.M. (1986) *Cell*, **47**, 995–1005.
Nasrin,N., Buggs,C., Fu Kong,X., Carnazza,J., Goebl,M. and Alexander-Bridges,M. (1991) *Nature*, **354**, 317–320.
Pontiggia,A., Whitfield,S., Goodfellow,P.N. and Bianchi,M.E. (1995) *Gene,* in press.
Sinclair,A.H. *et al.* (1990) *Nature*, **346**, 240–244.
Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendorff, J.W. (1991) *Methods Enzymol.*, **185**, 60–89.
Tabor,S. and Richardson,C.C. (1985) *Proc. Natl Acad. Sci. USA*, **82**, 1074–1078.
Thompson,J.F. and Landy,A. (1988) *Nucleic Acids Res.*, **16**, 9687–9705.
Tjian,R. and Maniatis,T. (1994) *Cell*, **77**, 5–8.
Travers,A.A. (1988) In Eckstein,F. and Lilley,D.M.J. (eds), *Nucleic Acids and Molecular Biology.* Springer Verlag, Berlin, Vol.2, pp.136–148.
Tucker,P.K. and Lundrigan,B.L. (1993) *Nature*, **364**, 715–717.
van de Wetering,M. and Clevers,H. (1992) *EMBO J.*, **11**, 3039–3044.
Whitfield,L.S.,Lovell-Badge,R. and Goodfellow,P.N. (1993) *Nature*, **364**, 713–715.
Wu,H.-M. and Crothers,D.M. (1984) *Nature*, **308**, 509–513.

Received on August 12, 1994