An SRY Mutation Causing Human Sex Reversal Resolves a General Mechanism of Structure-Specific DNA Recognition: Application to the Four-Way DNA Junction†

Richard Peters,‡,§ Chih-Yen King,‡ Etsuji Ukiyama,|| Sassan Falsafi,‡ Patricia K. Donahoe,|| and Michael A. Weiss*,‡,§,⊥

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts 02114, and Pediatric Surgical Research Laboratory and Department of Surgery, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114

Received November 3, 1994; Revised Manuscript Received January 19, 1995⊗

ABSTRACT: SRY, a genetic "master switch" for male development in mammals, exhibits two biochemical activities: sequence-specific recognition of duplex DNA and sequence-independent binding to the sharp angles of four-way DNA junctions. Here, we distinguish between these activities by analysis of a mutant SRY associated with human sex reversal (46, XY female with pure gonadal dysgenesis). The substitution (I68T in human SRY) alters a nonpolar side chain in the minor-groove DNA recognition α-helix of the HMG box [Haqq, C. M., King, C.-Y., Ukiyama, E., Haqq, T. N., Falsalfi, S., Donahoe, P. K., & Weiss, M. A. (1994) *Science 266*, 1494−1500]. The native (but not mutant) side chain inserts between specific base pairs in duplex DNA, interrupting base stacking at a site of induced DNA bending. Isotope-aided ^{1}H-NMR spectroscopy demonstrates that analogous side-chain insertion occurs on binding of SRY to a four-way junction, establishing a shared mechanism of sequence- and structure-specific DNA binding. Although the mutant DNA-binding domain exhibits >50-fold reduction in sequence-specific DNA recognition, near wild-type affinity for four-way junctions is retained. Our results (i) identify a shared SRY−DNA contact at a site of either induced or intrinsic DNA bending, (ii) demonstrate that this contact is not required to bind an intrinsically bent DNA target, and (iii) rationalize patterns of sequence conservation or diversity among HMG boxes. Clinical association of the I68T mutation with human sex reversal supports the hypothesis that specific DNA recognition by SRY is required for male sex determination.

Sexual dimorphism provides a model of a genetic switch between alternative programs of development (Lovell-Badge, 1993). The male phenotype in mammals is determined by SRY[1] (Sinclair *et al.*, 1990), a putative transcription factor (Cohen *et al.*, 1994) encoded by the short arm of the Y chromosome (Goodfellow & Lovell-Badge, 1993). Assignment of SRY as the testis-determining factor (TDF) is supported by studies of transgenic murine models (Koopman *et al.*, 1991) and human intersex abnormalities (Berta *et al.*, 1990; McElreavy *et al.*, 1992; Hawkins *et al.*, 1992ab; Harley *et al.*, 1992; Vilain *et al.*, 1993). Here, we resolve a general mechanism of structure-specific DNA recognition by analysis of a mutant SRY associated with human sex reversal (McElreavy *et al.*, 1992). Our results identify a contact between the protein and a four-way DNA junction. This contact, insertion of a nonpolar side chain between base pairs (Haqq *et al.*, 1994), is reminiscent of analogous (sequence-specific) contacts at sites of sharp DNA bending in crystal structures of the TATA-binding protein (Kim, Y., *et al.*, 1993; Kim, J. L., *et al.*, 1993) and prokaryotic transcription factor PurR (Schumacher *et al.*, 1994). Its extension to the sharp angles of a four-way DNA junction has implications for the evolution of HMG-box sequences and supports the hypothesis that the sex-determining function of SRY requires sequence-specific DNA recognition.

SRY contains a conserved motif (the HMG box) first described among high-mobility-group (HMG) nuclear proteins. This motif is also shared by a newly recognized class of transcription factors [Gubbay *et al.*, 1990; see Grosschedl *et al.* (1994) for review]. The HMG box exhibits a novel α-helical fold as determined by nuclear magnetic resonance (NMR) spectroscopy (Weir *et al.*, 1993; Read *et al.*, 1993; Jones *et al.*, 1994). The tertiary structure is L-shaped, defining convex and concave protein surfaces (Figure 1A). The HMG box exhibits an unusual range of DNA-binding activities. (i) Classical high-mobility group nuclear proteins are sequence nonspecific but bind to sites of altered DNA structure, such as supercoiled DNA, four-way DNA junctions, or intrinsically bent drug−DNA adducts (Pil &

† R.P. is a Howard Hughes Medical Institute Physician Postdoctoral Fellow. M.A.W. is an Established Investigator of the American Heart Association, Lucille P. Markey Scholar, and Bane Scholar at The University of Chicago. This work was supported in part by grants from the March of Dimes (M.A.W.), the NIH (HD30812 to P.K.D. and GM51558 to M.A.W.), and Reproductive Endocrine Sciences Center P30 (HD28138).

* To whom correspondence should be addressed.

‡ Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School.

§ Department of Medicine, Massachusetts General Hospital.

|| Pediatric Surgical Research Laboratory and Department of Surgery, Massachusetts General Hospital and Harvard Medical School.

⊥ Present address: Center for Molecular Oncology and Departments of Biochemistry and Molecular Biology and of Chemistry, The University of Chicago, Chicago, IL 60637-5419.

⊗ Abstract published in *Advance ACS Abstracts,* March 15, 1995.

[1] Abbreviations: BSA, bovine serum albumin; DG, distance geometry; HMG, high mobility group; HMQC, heteronuclear multiple-quantum coherence; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; PCR, polymerase chain reaction; RMSD, root-mean-square deviation; SA, simulated annealing; SRY, sex-determining region of the Y chromosome; TDF, testis-determining factor. Amino acids are designated by single-letter code.

0006-2960/95/0434-4569$09.00/0

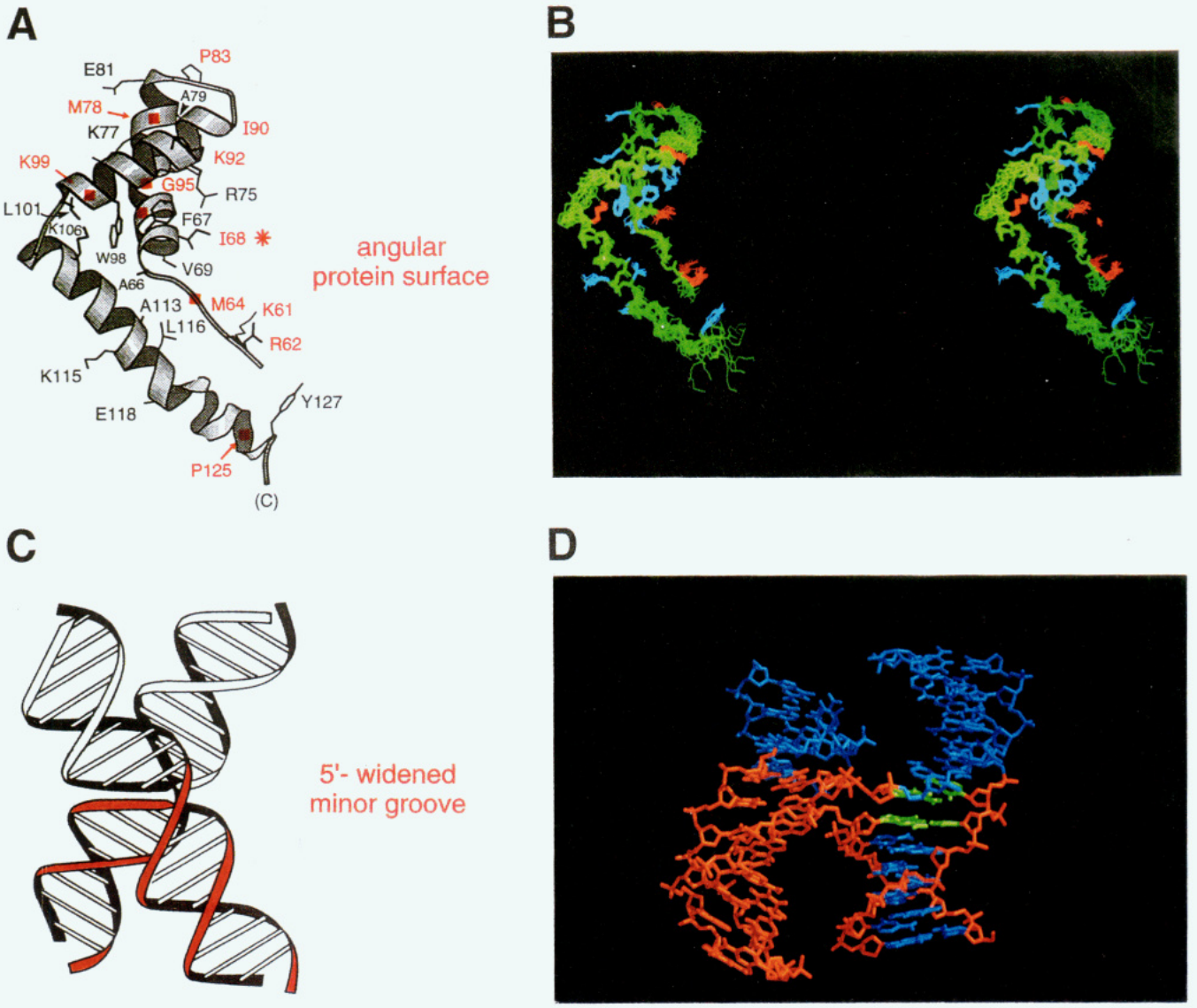


FIGURE 1: (A and B) Model of the human SRY HMG box (residues 61−128). In Molscript rendition (A) positions of mutations associated with sex reversal are shown in red; an asterisk indicates I68 (position 16 of the HMG box). In panel B (DG ensemble) are shown well-ordered side chains (RMSD < 2 Å); the arrow indicates I68. (C and D) Models of the four-way junction (von Kitzing *et al.,* 1990): panel C, schematic rendition of the Mg^{2+}-stabilized X structure (Lilley & Clegg, 1993); panel D, molecular mechanics model. The angle formed by divergent arms is shown in red; one set of junctional base pairs in (D) is shown in green.

Lippard, 1992). (ii) HMG-box transcription factors (including SRY) recognize specific DNA sequences (Nasrin *et al.,* 1991; Harley *et al.,* 1992; Haqq *et al.,* 1993) and induce sharp DNA bends (Van de Wetering & Clevers, 1992; Giese *et al.,* 1992). Unlike classical DNA-binding motifs (Sauer & Pabo, 1992), the HMG box binds primarily in the DNA minor groove (Van de Wetering & Clevers, 1992; Giese *et al.,* 1992). HMG-box transcription factors also exhibit sequence-independent binding to the sharp angles of four-way DNA junctions (Figure 1C; Ferrari *et al.,* 1992).

Does the HMG box exhibit a single mechanism of DNA binding, and if so, how can this mechanism account for diverse DNA-binding activities? What in particular is the relationship between sequence-specific and structure-specific DNA recognition? In the present study we compare how the SRY HMG box binds to a specific DNA duplex (an induced DNA bend) and nonspecific four-way DNA junction (an intrinsic DNA bend). Our results, based on isotope-assisted ^{1}H-NMR spectroscopy and site-directed mutagenesis, provide evidence for a shared mechanism of DNA binding. Remarkably, however, the two DNA-binding activities are distinguished by a mutation in the SRY HMG box associated with human sex reversal. The results provide insight into the biological role of SRY and the relationship between HMG box structure and function.

MATERIALS AND METHODS

Protein Purification. The SRY HMG box (85 residues) was expressed in *Escherichia coli* as a thrombin-cleavable fusion protein as described (Hinck *et al.,* 1993) with inclusion of a C-terminal His_6 tag, permitting affinity purification (Qian *et al.,* 1993). Final purification of the HMG-box fragment was accomplished by FPLC using a MonoS column (Pharmacia, Inc.). Purity was >98% as assessed by SDS−polyacrylamide gel electrophoresis. The His_6 tag does not affect protein folding (as determined by circular dichroism and ^{1}H-NMR) or specific DNA binding (as determined by gel-shift assay).

Isotopic Labeling. Selective biosynthetic incorporation of [^{13}C]isoleucine was accomplished in a prototrophic strain of

E. coli [BL21(DE3)] by feedback inhibition of endogenous biosynthesis with unlabeled valine and leucine in minimal medium. ^{13}C-HMQC spectra of the labeled protein (Griffey *et al.*, 1983) demonstrated >98% labeling with <2% metabolic diffusion of the label to other amino acids. Uniform ^{13}C and ^{15}N labeling was accomplished in the same strain using minimal medium containing [^{13}C]glucose and [^{15}N]ammonium sulfate.

Four-Way DNA Junctions. Sequences were obtained from previous studies. Individual duplex arms do not exhibit specific SRY binding. (i) The following four 30-base-pair oligonucleotides were prepared as described by Ferrari *et al.* (1992): *strand 1,* 5′-AGCGCTCTCACACGGCCCT-CCGCCCAGCTG; *strand 2,* 5′-CAGCTGGGCGGAGG-GGCGCGGACGTTAACC; *strand 3,* 5′-GGTTAAC-GTCCGCGCGGGTAATCTGGTAGA; and *strand 4,* 5′-TCTACCAGATTACCCCCGTGTGAGAGCGCT. The junctional bases are underlined. One strand was labeled with ^{32}P and annealed overnight with the other three strands from 80 to 4 °C in 50 mM Tris-HCl (pH 7.5), 5 mM dithiothreitol, 10 mM $MgCl_2$, and 50 mg/mL BSA. Efficiency of junction formation was >80% as estimated by 12% PAGE. The four-way junction was eluted from the gel; the recovered DNA was quantified using a ^{32}P scintillation counter. (ii) Junction J1, containing four 16-base-pair oligonucleotides, was prepared as described by Chazin and co-workers (Chen *et al.*, 1993): *strand 1,* 5′-CGCAATCCTGAGCACG-3′; *strand 2,* 5′-GCATTCGGACTATGGC-3′; 5′-GCCATAGTGGA-TTGCG-3′; and *strand 4,* 5′-CGTGCTCACCGAAT-GC-3′. The junctional bases are underlined. Strands (1 mM) were mixed in equimolar proportions in 10 mM potassium phosphate (pD 7.0), 50 mM NaCl, 0.05% NaN_3, and 5 mM $MgCl_2$, heated in an NMR tube to 90 °C, and annealed by slow cooling. The final ^{1}H-NMR spectrum was similar to that described (Chen *et al.*, 1993).

DNA-Binding Assay. Oligonucleotides were purchased from Oligos, Etc. (Wilsonville, OR). (i) The duplex probe was labeled with ^{32}P, annealed, and analyzed using the gel retardation assay as described (Haqq *et al.*, 1993; King & Weiss, 1993). Each reaction contained 50−500 nM protein and 0.8 nM labeled DNA in 10 mM potassium phosphate (pH 7.0), 50 ng/mL bovine serum albumin (BSA), 50 mM KCl, and 2.5 mM $MgCl_2$; the reaction was incubated for 1 h on ice. Only specific binding is observed under these conditions. (ii) For analysis of four-way DNA junctions, reactions contained 25−250 nM protein and 4 nM labeled junction in the same buffer.

Site-Directed Mutagenesis. Single amino acid substitutions were introduced into the SRY coding region in phage M13mp19RF by oligonucleotide-directed mutagenesis as described (Kunkel, 1985) and recloned by PCR into an expression plasmid (Hinck *et al.*, 1993). All constructions were verified by DNA sequencing. Purification of the variant proteins was as described for the wild type.

NMR Spectroscopy. Spectra were observed at 500 MHz using a Varian Unity spectrometer at Harvard Medical School. ^{1}H−^{13}C heteronuclear multiple-quantum coherence (HMQC) spectra were acquired by indirect detection as described (Griffey *et al.*, 1983).

Molecular Modeling. A model of a four-way DNA junction (von Kitzing *et al.*, 1990) was kindly provided by the authors. A model of the SRY HMG box was built from a homologous NMR structure (HMG1 box b; Weir *et al.*, 1993) by distance geometry and simulated annealing (DG/SA) by the method of Havel and Snow (1991) as described (Haqq *et al.*, 1994). In brief, 21 779 distance and 226 chirality restraints were used. Distance bounds were based on minimum and maximum distances in the NMR ensemble; 0.15 Å was added to the maximum and subtracted from the minimum when the total bound length was <3 Å. All interatomic distances, including protons, were included for identical residues. For conservative substitutions, distance constraints up to C_β and χ_1 and χ_2 dihedral restraints were also added. Ideal α-helical constraints (φ, ψ, and distances) were added to define secondary structure. Restraints were not introduced for nonconservative substitutions. There were no distance restraint violations. The SRY HMG box and HMG1 box b exhibit 25% sequence identity and 35% sequence similarity.

HYPOTHESIS

How can one protein motif exhibit both sequence-specific[2] and structure-specific DNA binding? The simplest hypothesis is *molecular mimicry:* that the DNA structure in a specific SRY complex (an induced DNA bend) resembles two arms of a four-way DNA junction (an intrinsic DNA bend; shown in red in Figure 1C and 1D). This hypothesis makes two testable predictions. (1) The HMG box employs a universal DNA-binding surface, complementary to a DNA bend whether induced or intrinsic. (2) Sequence- and structure-specific DNA recognition can be resolved by mutagenesis. In particular, side chains required to alter the structure of a specific DNA duplex may be dispensable in binding to a four-way junction. Analysis of phenotypes associated with such mutations would provide biological insight into the function of SRY in mammalian development.

MODEL

A specific complex between the SRY HMG box and duplex DNA exhibits a novel protein−DNA contact (King & Weiss, 1993). Analysis of nuclear Overhauser enhancements (NOEs) demonstrated that an isoleucine inserts between AT base pairs (arrow):

```
          ↓
5'-TGAT TGTTC-3'
3'-ACTA ACAAG-5'
```

The isoleucine spin system is shifted to high field, presumably by the ring currents of flanking DNA bases. The side chain enters through the DNA minor groove; only its tip (the δ-methyl group; resonance at −1.22 ppm indicated by an arrow and asterisk in Figure 2a) contacts internal thymidine N^3 imino protons. There are two isoleucine residues in the human SRY HMG box, I68[3] and I90 (Sinclair *et al.*, 1990; Whitfield *et al.*, 1994). ^{1}H-NMR resonances of the inserted isoleucine have been assigned by mutagenesis to I68 (position 16 in the HMG box motif; Haqq *et al.*, 1994). The three-dimensional environment of I68 has been inferred from homology modeling based on the solution structure of HMG1 box b (Figure 1A,B; Weir *et al.*, 1993; Haqq *et al.*, 1994). In this model (as in the parent structure) residue 68

[2] *Sequence-specific DNA binding* refers to recognition of specific sites in duplex DNA; *structure-specific DNA binding* designates recognition of noncanonical DNA structures (such as the sharp angles of a four-way junction) independent of base sequence.

[3] Residue numbers refer to the human SRY sequence (Whitfield *et al.*, 1993) unless otherwise indicated.

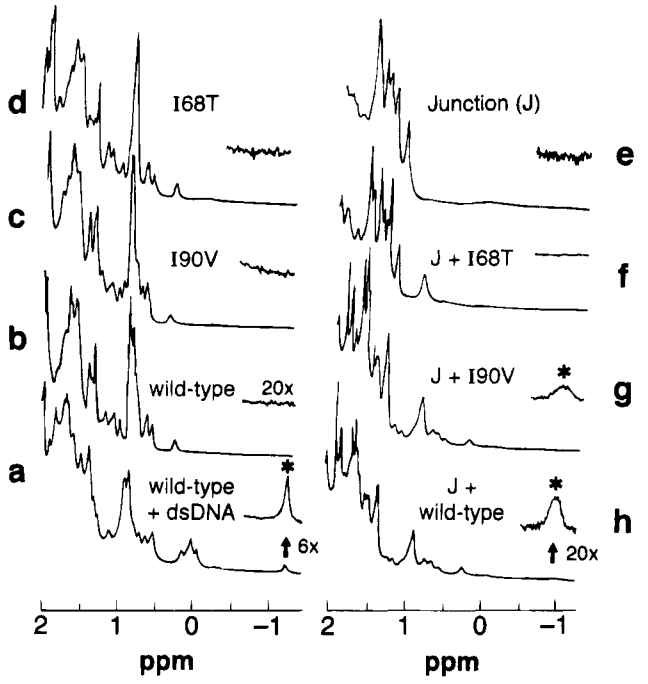


FIGURE 2: ^{1}H-NMR spectra of the (a) specific complex between the 15-bp target site and the native SRY HMG box (King & Weiss, 1993), (b) free native SRY domain, (c) I90V variant domain, (d) I68T variant domain, (e) free four-way junction J1 (Chen *et al.*, 1993), (f, g) variant SRY junctional complexes as labeled, and (h) wild-type SRY junctional complex. The upfield portion of each spectrum is enlarged as indicated. Asterisks and arrows indicate δ-CH_3 resonances of I68.

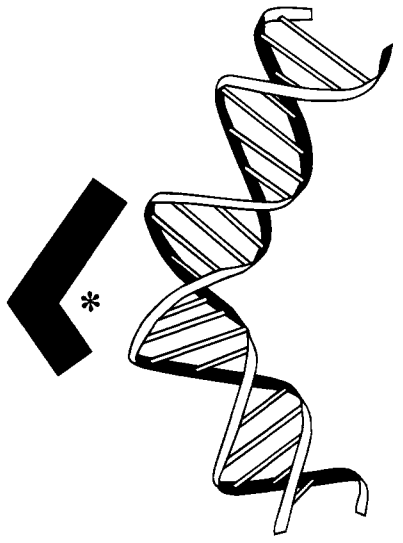


FIGURE 3: Schematic model of the L-shaped HMG box (left) and bent DNA site (right). The asterisk indicates the position of residue 16 of the HMG box as the crux of the concave protein surface ("inside the L"; Weir *et al.*, 1993; Read *et al.*, 1993; Jones *et al.*, 1994). Intermolecular NOEs indicate direct contact between residue 16 (I68 in human SRY; Whitfield *et al.*, 1993) and the DNA minor groove.

projects from α-helix 1 at the crux of an angular surface: the concave aspect of the L-shaped structure (Figure 3). Involvement of this surface in DNA binding is consistent with the inferred positions of mutations associated with human sex reversal (indicated in red in Figure 1A,B) and is supported by studies of homologous HMG-box proteins (Falciola *et al.*, 1994; Read *et al.*, 1994). Isoleucine insertion interrupts base stacking (but not base pairing) at the site of insertion (King & Weiss, 1993). This interaction is analogous to features of the TATA-binding protein (TBP)−DNA complex (Kim *et al.*, 1993a,b) and would likewise be expected to induce a sharp DNA bend as demonstrated by anomalous electrophoretic mobility (Van de Wetering & Clevers, 1992; Giese *et al.*, 1992). We would also expect that the DNA minor groove must widen to accommodate α-helix 1 (Haqq *et al.*, 1994); analogous widening occurs in the PurR complex (Schumacher *et al.*, 1994) and in the TBP−DNA complex to accommodate a β-sheet (Kim, Y., *et al.*, 1993; Kim, J. L., *et al.*, 1993).

RESULTS

Prediction 1: An Analogous Ile Contact Occurs in a Four-Way DNA Junction. The mechanism of SRY binding to four-way junction J1 with Mg^{2+} (Chen *et al.*, 1993) was investigated by ^{1}H-NMR. Strikingly, the spectrum of a 1:1 complex contains a methyl resonance shifted to high field (−1.03 ppm; asterisk in Figure 2h), which is similar to that of I68-δCH_3 in the spectrum of a specific complex (Figure 2a). Such a correspondence, if verified by resonance assignment, would demonstrate use of a similar or overlapping DNA-binding surface (prediction 1). Resonance assignment was obtained in two steps by isotopic labeling (Figure 4) and site-directed mutagenesis (Figure 2e−h). (i) Selective labeling of the SRY domain with ^{13}C-enriched isoleucine is shown by comparison of HMQC spectra (Figure 4A,B). The label confers the expected 140-Hz $^1J_{CH}$ coupling to the upfield resonance (spectrum a in Figure 4C), verifying its classification as an isoleucine resonance. (ii) Assignment to I68 is demonstrated by inspection of the spectra of variant junctional complexes I68T and I90V (Figure 2f,g). The site of I68 insertion in the four-way junction has not been determined. Since the SRY HMG box does not bind to the isolated arms (data not shown), the insertion site is presumed to be at or near the junctional base pairs (green in Figure 1D). Since in the presence of Mg^{2+} the four-way DNA junction contains two acute and two obtuse angles (Lilley & Clegg, 1993), it is likely that the 1:1 complex represents an equilibrium among two or more binding configurations. Such an equilibrium could account for the large line width of the I68 methyl resonance (relative to thymidine CH_3 resonances in the DNA) either by intermediate exchange or by partial overlap of inequivalent I68 resonances in slow exchange. Distinct 1:1 complexes are not resolved by the gel mobility shift assay (Ferrari *et al.*, 1992).

Prediction 2: Isoleucine Insertion Is Required for Specific DNA Recognition but Is Dispensable for Four-Way Junction Binding. We test prediction 2 by analysis of a variant SRY HMG box I68T, containing a substitution associated with human sex reversal (McElreavy *et al.*, 1992). A conservative substitution of the other isoleucine (I90V, expected to pack in the hydrophobic core; Weir *et al.*, 1993; Read *et al.*, 1993; Jones *et al.*, 1994; Haqq *et al.*, 1994) was constructed as a control. ^{1}H-NMR spectra (Figure 2b−d) and thermodynamic stabilities (not shown) of the variant and native domains are similar.

Threonine, because of its shorter length (lacking a δ-substituent) and polar substituent (β-OH function), would not be expected to penetrate between base pairs. If side-chain insertion stabilizes an induced DNA bend and if such bending is required to bind a specific duplex site, then the I68T

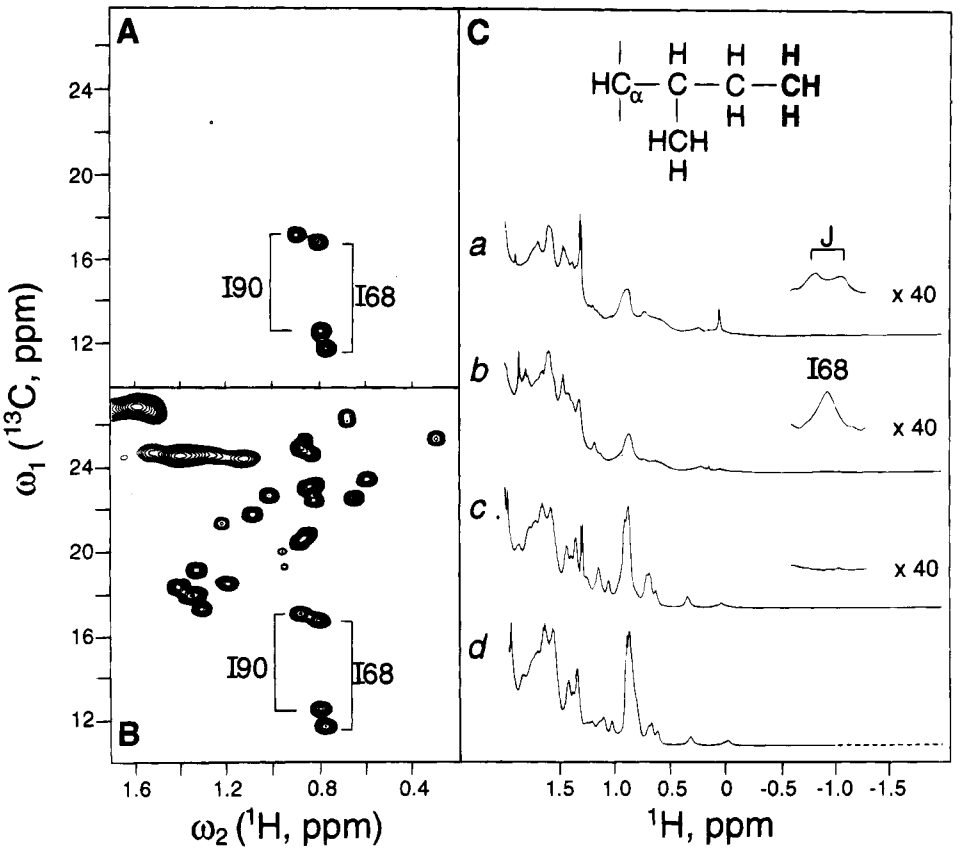


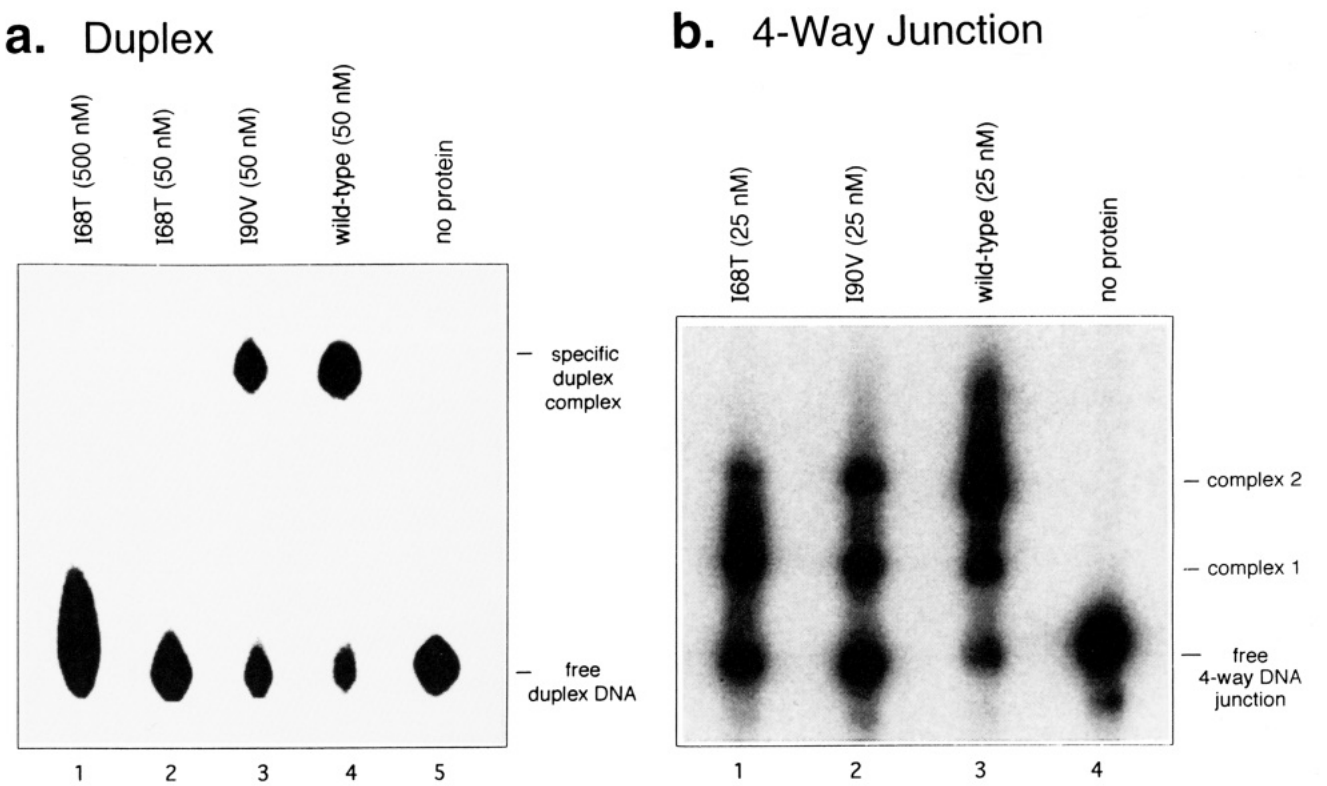
FIGURE 4: (A) HMQC spectrum of the SRY domain selectively labeled with [^{13}C]isoleucine. (B) HMQC spectrum of the SRY domain uniformly enriched with ^{13}C and ^{15}N. γ'-CH_3 and δ-CH_3 resonances (near 17 and 12 ppm in the ^{13}C dimension, respectively) are outlined. (C) Upper panel: structure of the isoleucine side chain (δ-CH_3 in boldface). Lower panel: ^{1}H-NMR spectra of the (a) junctional complex with a selective isoleucine ^{13}C label, (b) junctional complex without a label, (c) labeled domain alone, and (d) unlabeled domain alone.

domain would be expected to exhibit reduced specific binding. In contrast, binding to a Mg^{2+}-stabilized four-way DNA junction might be maintained, since molecular mechanics calculations (von Kitzing *et al.*, 1990) predict that its structure (Figure 1C,D) exhibits the three postulated hallmarks of a specific duplex complex: (a) the arms define sharp DNA bends with intact base pairing; (b) junctional base pairs are displaced at the center of the intrinsic bend, mimicking displacement by an inserted amino acid; and (c) the minor groove is widened 5′ to the site of strand exchange, as would be required to accommodate an α-helix. Although the structure of a four-way junction has not to date been determined by X-ray crystallography, this model is supported by biochemical and spectroscopic studies [reviewed by Lilley and Clegg (1993)].

Binding of native and variant domains to a specific duplex DNA site was tested by gel retardation (Figure 5a). Whereas native and I90V domains each bind with an apparent dissociation constant of 25 nM (lanes 4 and 3), binding of the I68T variant is not detectable (lanes 1 and 2). The assay's sensitivity implies that specific binding is weakened by at least 100-fold. In contrast, the I68T domain exhibits almost native binding to four-way junctions (Figure 5b). Such binding also occurs at nanomolar protein concentrations and gives rise to two complexes, presumably representing successive binding to distinct junctional angles (Ferrari *et al.*, 1992). The isolated arms exhibit no binding to the native or variant domains. Similar results are obtained with a variety of junctional sequences (not shown). Preserved binding of the I68T domain to four-way junctions is in accord with the diversity of side-chain functionalities (length, charge, and polarity) at this position among sequence-nonspecific HMG-box proteins (I, M, F, A, E, N, and Q; box in Figure 5d). In contrast, sequence-specific HMG boxes contain only isoleucine, methionine, and phenylalanine (Figure 5c). We speculate that the latter conservation is imposed by specific steric and polarity requirements of side-chain–DNA insertion and resultant protein-induced DNA bending.

DISCUSSION

Identification of SRY as the sex-determining locus of the human Y chromosome was accomplished by "reverse genetics": analysis of deletions and translocations associated with phenotypic sex reversal (XY females and XX males) [Sinclair *et al.*, 1990; for review, see Goodfellow and Lovell-Badge (1993)]. Its genetic function as the long-sought testis-determining factor (TDF) was verified in transgenic XX mice (Koopman *et al.*, 1991). Because this function was assigned in the absence of biochemical characterization, however, SRY's mechanism of action poses an unsolved problem. An initial clue was provided by observation that SRY contains a newly recognized DNA-binding motif, the HMG box [Figure 1A,B; reviewed by Grosschedl *et al.* (1994)]. This motif is conserved among eukaryotes and defines an ancestral gene family (Nerr, 1992; Laudet *et al.*, 1993). The SRY HMG box binds with similar affinity to specific DNA sequences (Nasrin *et al.*, 1992; Harley *et al.*, 1992; Haqq *et al.*, 1993; Cohen *et al.*, 1994; Hawley *et al.*, 1994) and the sharp angles of four-way DNA junctions (Ferrari *et al.*, 1992). A model of a Mg^{2+}-stabilized four-way junction is shown in Figure 1C,D (von Kitzing *et al.*, 1990). The sharp angles, partially unstacked central base pairs, and widened



c. Sequence-Specific

hSRY	RVKRPMNAF	I	VWSRDQRRKMALENP
mSRY	HVKRPMNAF	M	VWSRGERHKLAQQNP
sSRY	AF	I	VWSRERRRKVALENP
rSRY	AF	M	VWSQHQRRQVALENP
smSRY	AF	M	VWSQTQRRQVALQNP
maSRY	AF	M	IWSRSQRRQVALQNP
dSRY	RVKRPMNAF	M	VWSQTQRRQVALQNP
gSRY	RVKRPMNAF	I	VWSRDQRRKMALENP
bSRY	RVKRPMNAF	I	VWSRDQRRKMALENP
oSRY	RVKRPMNAF	I	VWSRDQRRKMALENP
marSRY	RVKRPMNAF	I	VWSRDQRRKMAVENP
chSRY	RVKRPMNAF	F	VWSRDQRRKMALENP
p.chSRY	RVKRPMNAF	F	VWSRDQRRKMALENP
Sox1,2,3	MNAF	M	VWSRGQRRKMALENP
Sox4	MNAF	M	VWSQIERRKIMEQSP
LEF-1	HIKKPLNAF	M	LYMKEMRANVVAECT
TCF-1	TIKKPLNAF	M	LYMKEMRAKVIAECT
ROX-1	KIPAPKNAF	I	LFRQHYHRILIDEWT
STE-11	SVKRPLNSF	M	LYRRDRQAEI----P
MAT-A1	KIPRPPNAY	I	LYRKDHHREIREQNP
MAT-MC	RTPRPPNAF	I	LYRKEKHATLLKSNP
IRE-ABP	HIKRPNMAF	M	VWSQIERRKIMEQSP

putative α-helix 1

d. Structure-Specific

SSRP-1	APKRPMSAY	M	LWLNASREKIKSDHP
T-160	APKRPMSAY	M	LWLNASREKIKSDHP
ABF-2.1	GPKRPTSAY	F	LYLQDHRSQFVKENP
ABF-2.2	PPKKPAGPF	I	KYANEVRSQVFAQHP
HMG-D	KPKRPLSAY	M	LWLNSARESIKRENP
HMG-T.1	KPRGKMSSY	A	YFVQTRREEHKKKHP
HMG-T.2	APKRPSSAF	F	IFCADFRPQVKGETP
rHMG-1.A	KPRGKMSSY	A	FFVQTCREEHKKKHP
rHMG-1.B	APKRPPSAF	F	LFCSEYRPKIKGEHP
pHMG-2.A	KPRGKMSSY	A	FFVQYCREEHKKKHP
pHMG-2.B	APKRPPSAF	F	LFCSEHRPKIKSEHP
NHP6A	APKRALSAY	M	FFANENRDIVRSENP
mTF1.1	CPKKPVSSY	L	RFSKEQLPIFKAQNP
mTF1.2	KPKRPRSAY	N	VYVAERFQEAKGDSP
hUBF.1	FPKKPLTPY	F	RFFMEKRAKYAKLHP
hUBF.2	IPEKPKTPQ	Q	LWYTHEKKVYLKVRP
hUBF.3	KPKRPVSAM	F	IFSEEKRRQLKEERP
hUBF.4	LPESPKRAE	E	IWQQSVIGDYLARFK

α-helix 1

FIGURE 5: (Panels a and b) Gel retardation assays comparing binding to a specific DNA duplex (panel a, 5′-GGGGTG**TTTGTG**CAG-3′) and binding to a four-way junction (panel b). The duplex sequence is derived from a putative DNA control site in the promoter of the Müllerian inhibiting substance *(MIS)* gene (Haqq *et al.,* 1993); similar results have been obtained with an optimal SRY-binding site with the central sequence 5′-ATTGTT (Haqq *et al.*, 1994). The junction sequence (Materials and Methods) is as described by Ferrari *et al.* (1992); analogous results are obtained with the junction sequence J1 (Chen *et al.,* 1993). (Panels c and d) Partial sequences of sequence-specific and structure-specific HMG boxes (Ner, 1992; Laudet *et al.,* 1993; Whitfield *et al.,* 1993; Grosschedl *et al.,* 1994). In each case the position corresponding to I68 in human SRY is boxed. Abbreviations: hSRY, human; mSRY, mouse; sSRY, sheep and cattle; rSRY, rabbit; smSRY, sminthopsis; maSRY, macropus; dSRY, dunnart; gSRY, gorilla; bSRY, baboon; oSRY, orangutan; marSRY, marmoset; chSRY, chimpanzee; p.chSRY, pyg-chimp; Sox, SRY box gene; LEF-1, lymphocyte enhancer factor-1; TCF-1, human T cell-specific transcription factor; ROX-1, repressor of the hypoxic genes of *Saccharomyces cerevisiae*; STE-11, sexual development of *Saccharomyces pombe*; MAT-A1, mating for type recognition A-1 phenotype; MAT-MC, mating-type protein of *S. pombe*; IRE-ABP, insulin-response element binding; SSRP-1, human structure-specific recognition protein; T-160, murine recombination signal sequence binding protein; ABF, yeast ARS-binding protein; HMG-D, *Drosophila* HMG; HMG-T, trout HMG; rHMG, rat HMG; pHMG, pig HMG-1; NHP6A, non-histone chromosomal protein from *S. cerevisiae*; mTF1, mitochondrial transcription factor; hUBF, RNA polymerase I transcription factors.

minor groove of the four-way DNA junction (Lilley & Clegg, 1993) are likely to mimic characteristic features of a specific SRY−DNA complex (Ferrari *et al.*, 1992; King & Weiss, 1993).

Binding to four-way junctions provides a model for structure-specific but sequence-independent DNA recognition. Such binding can in principle provide a mechanism of genetic regulation. *In vitro*, HMG-1 and -2 , which lack sequence specificity, can stimulate transcriptional initiation by RNA polymerases II and III (Johns, 1982; Tremethick & Molloy, 1986). Although the mechanism of activation is not well characterized, HMG binding to (or stabilization of) structurally altered promoter DNA has been proposed to regulate assembly of an initiation complex. HMG-box proteins UBF (an RNA polymerase I transcription factor) and mTF1 (a mitochondrial polymerase transcription factor) bind to specific positions relative to their respective transcriptional start sites, yet neither recognizes a clear consensus DNA sequence (Fisher *et al.*, 1989; Bell *et al.*, 1989). Such studies indicate that transcriptional regulation by HMG-box proteins can be mediated by recognition of distinct topological domains rather than by recognition of duplex target sequences. Studies of phage λ recombination complexes demonstrate an analogous role for sequence-independent but structure-specific DNA-binding proteins, permitting interchangeability of integration host factor (IHF) with eukaryotic HMG1 and HMG2 proteins (Segall *et al.*, 1994). The respective contributions of sequence- and structure-specific recognition to the biological function of SRY or homologous HMG-box transcription factors have not been defined. In particular, altered DNA structures in the preinitiation complex (including the initiator element) may provide targets with features of both sequence- and structure-specific binding.

In this paper we have distinguished between the sequence- and structure-specific DNA-binding activities of SRY by analysis of a mutant protein that selectively retains one activity but not the other. This mutation (I68T) occurs at a DNA contact site (King & Weiss, 1993; Haqq *et al.*, 1994) and is associated with human sex reversal (McElreavy *et al.*, 1992; Vilain *et al.*, 1993). Because the structure of SRY has not been determined, experimental design is based on the NMR structure of an HMG-1 fragment (box B; Weir *et al.*, 1993; Read *et al.*, 1993): three α-helices in a novel L-shaped fold (Figure 1A,B). A homology model of the SRY HMG box has been constructed by distance geometry by the method of Havel and Snow (1991) as previously described (Haqq *et al.*, 1994). Its angular surface is proposed to be complementary to the structure of an induced or intrinsic DNA bend. The inferred position of I68 (asterisk in Figure 3) identifies the concave surface ("inside the L") as in contact with DNA. This hypothesis is motivated by the overall shape of the HMG box and not by the uncertain details of the homology model. The dramatic upfield ^{1}H-NMR shift of I68 in both specific and junctional complexes strongly suggests a shared mechanism of DNA binding. We speculate that I68 represents a class of side chains required to induce specific DNA bending but not required to bind an intrinsically bent target. Insertion of a nonpolar (or aromatic) side chain between base pairs, proposed as a mechanism of specific DNA bending, is also observed in the bent PurR- and TATA-binding protein (TBP)−DNA complexes (Kim, Y., *et al.*, 1993; Kim, J. L., *et al.*, 1993; Schumacher *et al.*, 1994) and is reminiscent of features of protein−RNA recognition (Fedor, 1994).

How SRY directs testicular differentiation—and in particular, how site-specific DNA bending contributes to the regulation of gene expression—are not known. The clinical correlation of the I68T mutation in SRY with human sex reversal supports the hypothesis that sequence specificity is required (Nasrin *et al.*, 1991; Harley *et al.*, 1992; Haqq *et al.*, 1993, 1994) whereas binding to four-way junctions is insufficient. Because the four-way junction is not fully representative of the range of altered DNA structures in chromatin, the present results do not exclude other forms of structure-specific DNA recognition, as might occur in a preinitiation complex. Further progress in understanding the biological role of SRY−DNA interactions will require delineation of SRY-regulated genes (Lovell-Badge, 1994). Characterization of a downstream pathway of gene expression (Haqq *et al.*, 1994; Foster *et al.*, 1994; Wagner *et al.*, 1994) will provide an opportunity to dissect the role of directed DNA bending in a developmental switch.

ADDED IN PROOF

T-Cell factor-1 (Tcf-1), an HMG-box protein proposed to act as an architectural element to organize the spatial structure of T-cell-specific enhancers, has recently been shown to be required for thymocyte differentiation (Verbeek *et al.*, 1995). Because Tcf-1 and SRY exhibit similar DNA-binding properties and atypical transcriptional regulatory activity in cotransfection assays, similar biochemical mechanisms may control developmental decisions in the two systems.

ACKNOWLEDGMENT

We thank J. P. Lee, E. Rivera, and M. Lachenmann for assistance with NMR measurements; F. Tao for thermodynamic stability measurements; T. F. Havel for the program DGII and advice; C. Haqq, M. Kurian, and G. Warneck for assistance with site-directed mutagenesis; D. C. Page for SRY cDNA; A. Hinck and J. Markley for the expression plasmid pTSN; D. C. Wiley for use of the CD spectropolarimeter; W. Chazin, J. Habener, H. Keutmann, H. Kronenberg, H. Nash, and K. Struhl for helpful discussion; and S. Diekmann and E. von Kitzing for kindly providing four-way junction coordinates.

REFERENCES

Bell, S. P., Pikaard, C. S., Reeder, R. H., & Tjian, R. (1989) *Cell 59*, 489−497.

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

Bianchi, M. E., Beltrame, M., & Paonessa, G. (1989) *Science 243*, 1056−1058.

Chen, S. M., Heffron, F., & Chazin, W. J. (1993) *Biochemistry 32*, 319−326.

Cohen, D. R., Sinclair, A. H., & McGovern, J. D. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 4372−4376.

Du, W., Thanos, D., & Maniatis, T. (1993) *Cell 74*, 887−898.

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

Fedor, M. N. (1994) *Nature Struct. Biol. 1*, 267−269.

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

Fisher, R. P., Parisi, M. A., & Clayton, D. A. (1989) *Genes Dev. 3*, 2202−2217.

Foster, J. W., Dominguez-Steglich, M. A., Guioli, S., Kwok, C., Weller, P. A., Stevanovic, M., Weissenback, J. D., Mansour, S., Young, I. D., Goodfellow, P. N., Brook, J. D., & Schafer, A. J. (1994) *Nature 372*, 525−529.
Giese, G., Cox, J., & Grosschedl, R. (1992) *Cell 69*, 185−195.
Giese, K., Pagel, J., & Grosschedl, R. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 3368−3372.
Goodfellow, P. N., & Lovell-Badge, R. (1993) *Annu. Rev. Genet. 27*, 71−92.
Griffey, R. H., Poulter, C. D., Bax, A., Hawkins, B. L., Yamaizumi, Z., & Nishimura, S. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 5895−5897.
Grosschedl, R., Giese, K., & Pagel, J. (1994) *Trends Genet. 10*, 94−100.
Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Münsterberg, A., Goodfellow, P. N., & Lovell-Badge, R. (1990) *Nature 346*, 245−250.
Haqq, C. M., King, C.-Y., Donahoe, P. K., & Weiss, M. A. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 1097−1101.
Haqq, C. M., King, C.-Y., Ukiyama, E., Haqq, T. N., Falsalfi, S., Donahoe, P. K., & Weiss, M. A. (1994) *Science 266*, 1494−1500.
Harley, V. R., Jackson, D. I., Hextall, P. J., Hawkins, J. R., Berkovitz, G. D., Sockanathan, S., Lovell-Badge, R., & Goodfellow, P. N. (1992) *Science 255*, 453−456.
Harley, V. R., Lovell-Badge, R., & Goodfellow, P. N. (1994) *Nucleic Acids Res. 22*, 1500−1501.
Havel, T. F., & Snow, M. E. (1991) *J. Mol. Biol. 217*, 1−7.
Hawkins, J. R., Taylor, A., Berta, P., Levilliers, J., Van der Auwera, B., & Goodfellow, P. N. (1992a) *Hum. Genet. 88*, 471−474.
Hawkins, J. R., Taylor, A., Goodfellow, P. N., Migeon, C. J., Smith, K. D., & Berkovitz, G. D. (1992b) *Am. J. Hum. Genet. 51*, 979−984.
Hinck, A. P., Walkenhorst, W. F., Westler, W. M., Choe, S., & Markley, J. L. (1993) *Protein Eng. 6*, 221−227.
Holley, L. H., & Karplus, M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 152−156.
Johns, E. W. (1982) *The HMG Chromosomal Proteins*, Academic Press, London, U.K.
Jones, D. N. M., Searles, M. A., Shaw, G. L., Churchill, M. E. A., Ner, S. S., Keeler, J., Travers, A. A., & Neuhaus, D. (1994) *Structure 2*, 609−627.
Kim, J. L., Nikolov, D. B., & Burley, S. K. (1993) *Nature 365*, 520−527.
Kim, Y., Geiger, J. H., Hahn, S., & Sigler, P. B. (1993) *Nature 365*, 512−519.
King, C.-Y., & Weiss, M. A. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 11990−11994.
Klug, A. (1993) *Nature 365*, 486−487.
Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P. N., & Lovell-Badge, R. (1991) *Nature 351*, 117−121.
Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 488−492.
Laudet, V., Stehelin, D., & Clevers, H. (1993) *Nucleic Acids Res. 21*, 2493−2501.
Lilley, D. M., & Clegg, R. M. (1993) *Annu. Rev. Biophys. Biomol. Struct. 22*, 299−328.
McElreavy, K., Vilain, E., Abbas, N., Costa, J. M., Souleyreau, N., Kucheria, K., Boucekkine, C., Thibaud, E., Brauner, R., & Flamant, F. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 11016−11020.
Nasrin, N., Buggs, C., Kong, X. F., Carnazza, J., Goebl, M., & Alexander Bridges, M. (1991) *Nature 354*, 317−320.
Ner, S. S. (1992) *Curr. Biol. 2*, 208−210.
Pabo, C. O., & Sauer, R. T. (1992) *Annu. Rev. Biochem. 61*, 1053−1095.
Pil, P. M., & Lippard, S. J. (1992) *Science 256*, 234−237.
Qian, X., Gozani, S. N., Yoon, H.-S., Jeon, C., Agarwal, K., & Weiss, M. A. (1993) *Biochemistry 32*, 9944−9959.
Read, C. M., Cary, P. D., Crane-Robinson, C., Driscoll, P. C., & Norman, D. G. (1993) *Nucleic Acids Res. 21*, 3427−3436.
Read, C. M., Cary, P. D., Preston, N. S., Lnenicekalen, M., Cary, P. D., & Crane-Robinson, C. (1994) *EMBO J. 13*, 5639−5646.
Schumacher, M. A., Choi, K. Y., Zalkin, H., & Brennan (1994) *Science 266*, 763−772.
Segall, A. N., Goodman, S. D., & Nash, H. A. (1994) *EMBO J. 13*, 4536−4548.
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., & Goodfellow P. N. (1990) *Nature 346*, 240−244.
Struhl, K. (1994) *Science 263*, 1103−1104.
Tremethick, D. J., & Molloy, P. L. (1986) *J. Biol. Chem. 261*, 6986−6992.
Van de Wetering, M., & Clevers, H. (1992) *EMBO J. 11*, 3039−3044.
Verbeek, S., Izon, D., Hofhuis, F., Robanus-Maandag, E., te Riele, H., Van de Wetering, M., Oosterwegel, M., Wilson, A., MacDonald, H. R., & Clevers, H. (1995) *Nature 374*, 70−74.
Vilain, E., Jaubert, F., Fellous, M., & McElreavey, K. (1993) *Differentiation 52*, 151−159.
Vivian, N., Goodfellow, P. N., & Lovell-Badge, R. (1990) *Nature 346*, 245−250.
von Kitzing, E., Lilley, D. M. J., & Diekmann, S. (1990) *Nucleic Acids Res. 18*, 2671−2683.
Wagner, T., Wirth, J., Meyer, J., Zabel, B., Held, M., Zimmer, J., Pasantes, J., Bricarelli, F. D., Keutel, J., Hustert, E., Wolf, U., Tommerup, N., Schempp, W., & Scherer, G. (1994) *Cell 79*, 1111−1120.
Weir, H. M., Kraulis, P. J., Hill, C. S., Raine, A. R. C., Laue, E. D., & Thomas, J. O. (1993) *EMBO J. 12*, 1311−1319.
Whitfield, L. S., Lovell-Badge, R., & Goodfellow, P. N. (1993) *Nature 364*, 713−715.

BI9425765