

# Floppy SOX: Mutual Induced Fit in HMG (High-Mobility Group) Box-DNA Recognition

Michael A. Weiss

Department of Biochemistry  
Case Western Reserve University  
Cleveland, Ohio 44106

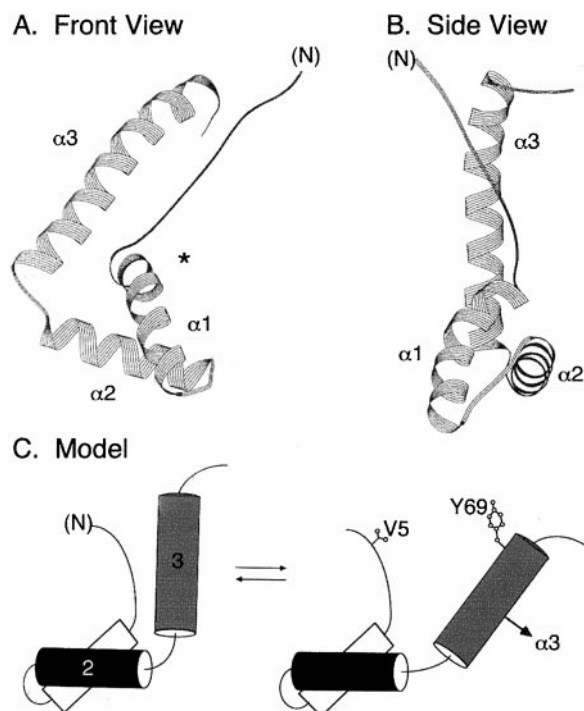
**The high-mobility group (HMG) box defines a DNA-bending motif of broad interest in relation to human development and disease. Major and minor wings of an L-shaped structure provide a template for DNA bending. As in the TATA-binding protein and a diverse family of factors, insertion of one or more side chains between base pairs induces a DNA kink. The HMG box binds in the DNA minor groove and may be specific for DNA sequence or distorted DNA architecture. Whereas the angular structures of non-sequence-specific domains are well ordered, free SRY and related autosomal SOX domains are in part disordered. Observations suggesting that the minor wing lacks a fixed tertiary structure motivate the hypothesis that DNA bending and stabilization of protein structure define a coupled process. We further propose that mutual induced fit in SOX-DNA recognition underlies the sequence dependence of DNA bending and enables the induction of promoter-specific architectures. (Molecular Endocrinology 15: 353–362, 2001)**

## INTRODUCTION

The high-mobility group (HMG) box defines a superfamily of eukaryotic DNA-binding proteins of central importance in mammalian gene regulation (1). This approximately 80-residue domain, originally described in abundant nonhistone chromosomal proteins HMG1 and HMG2, exhibits an unusual L-shaped structure (Fig. 1A and Ref. 2). Three  $\alpha$ -helices and an N-terminal  $\beta$ -strand pack to form major and minor wings (3–6). Each wing is formed by distinct elements of secondary structure. The major wing comprises  $\alpha$ -helix 1,  $\alpha$ -helix

2, the first turn of  $\alpha$ -helix 3, and connecting loops; the minor wing comprises the N-terminal  $\beta$ -strand, the remainder of  $\alpha$ -helix 3, and the C-terminal segment. Together, the L-shaped structure presents an angular inner surface as a template for DNA bending (*asterisk* in Fig. 1A; Refs. 7–11). A side view of the HMG box illustrates its flat architecture (Fig. 1B). Two groups of HMG boxes are distinguished by their DNA-binding properties. Whereas HMG1 and related proteins typically contain two or more HMG boxes that recognize distorted DNA structures with weak or absent sequence specificity, specific architectural transcription factors contain one HMG box that recognizes both distorted DNA structures and specific DNA sequences (2). Each group docks within a widened minor groove, directs the sharp bending of an underwound double helix, and can enhance binding of unrelated DNA-binding motifs to neighboring DNA sites (7–11). The extent of DNA bending varies among HMG boxes but in each case the protein binds on the outside of the DNA bend to compress the major groove. The dramatic effects of HMG boxes on DNA structure are proposed to contribute to the assembly of specific transcriptional preinitiation complexes and, in turn, to the regulation of gene expression (12, 13).

This minireview focuses on the unusual conformational repertoire of Sox proteins (14, 15), a subgroup of specific HMG-box factors defined by similarity to Sry (16), the mammalian testis-determining factor encoded by the Y chromosome (17). Designated Sox in relation to the Sry box, this subgroup is ubiquitous in the animal kingdom and involved in diverse developmental processes, including germ layer formation, cell type specification, and organogenesis (14, 15). More than 20 Sox genes have been identified based on greater than 50% sequence identity with the HMG box of Sry. Genetic analyses of Sox genes in humans, mice, and *Drosophila melanogaster* have demonstrated essential roles in specific cell fate decisions (18–21). Mutations or deletions in human SRY are a cause of Swyer's syndrome, in which failure of testic-



**Fig. 1.** Architectural Elements of the HMG Box

A and B, Ribbon models of a nonspecific HMG box (3–6) showing front view (A) and side view (B).  $\alpha$ -Helices 1 and 2 form the major wing; helix 3 and the N-terminal  $\beta$ -strand form the minor wing. Asterisk indicates position of hydrophobic wedge at crux of an angular protein surface. C, Proposed model of a specific HMG box in which the minor wing exists in equilibrium as an ensemble of open and closed structures (see Fig. 3D). Val5 and Y69 (right panel) provide NMR markers of the structure of the minor wing (see upper panel of Fig. 2). In a specific complex an interface forms between helix 3 and the N-terminal  $\beta$ -strand (asterisk at left).

ular differentiation in a 46,XY embryo leads to a female somatic phenotype and sterility (17). XY sex reversal can also occur with variable penetrance due to mutations in *SOX9*, a gene on human chromosome 19 (18, 19). Such mutations cause campomelic dysplasia, a syndrome of bony abnormalities associated with XY gonadal dysgenesis. Whereas clinical mutations in *SRY* cluster in its HMG box (7), *SOX9* mutations are widely distributed in its coding region (22).

Sox genes are classified in seven families (designated A–G) based on extent of homology (>80% within a family). The families exhibit similar DNA-binding and DNA-bending properties (14, 15). Random binding site selection *in vitro* has revealed a shared specificity for a core consensus sequence (5′-ACAAT-3′; Table 1) with subtle distinctions in preferences for flanking nucleotides (23–26). Although such chemical specificity is less stringent than that of classical major-groove DNA-binding motifs, functional specificity is enhanced by lineage-specific gene expression. The sequence specificity of Sox-9 and Sox-10 can also be made more stringent by cooperative binding of protein dimers to neighboring DNA target sites (27). Dimeriza-

**Table 1.** Consensus SRY- and SOX Binding Sites Defined by Random Binding-Site Selection<sup>a</sup>

HMG box	Sequence	Ref.
SRY	5′- <b>Tt</b> /aGAACAAT <b>AG</b> -3′	(23)
Sox5	5′- <b>Tt</b> /aAACAATA-3′	(24)
Sox17	5′- <b>AG</b> AACAAT <b>GG</b> -3′	(25)
Sox9	5′- <b>AG</b> AACAAT <b>G</b> -3′	(26)

<sup>a</sup> Differences among flanking oligonucleotides appear in bold; small letters t/a indicate that either T or A is tolerated at this position.

tion is DNA dependent and mediated by a conserved N-terminal protein segment. Although the structural basis of cooperativity is not understood, analysis of the interaction of Sox10 with a Sox10 response element active in neural crest-derived lineages [the protein zero *P(0)* gene] has shown that DNA-dependent dimerization markedly enhances specific DNA affinity and extent of induced DNA bending (27). It is not known whether changes in DNA structure can, by themselves, contribute to cooperativity, *i.e.* independently of putative DNA-dependent protein-protein interactions. It is possible, for example, that initial DNA bending and unwinding induced by one HMG box can facilitate binding of a second HMG box to an adjoining DNA site.

Specific DNA bends (typically in the range 70–90°) may disallow or facilitate DNA binding by unrelated transcription factors (*i.e.* proteins of other structural classes) to adjoining DNA sites or facilitate protein-protein interactions by flanking DNA-bound factors. HMG box-induced DNA bends may also recruit trithorax or polycomb group proteins, proposed to alter the higher-order chromatin structure leading to the long-range regulation of gene expression. The identification of target sites for Sox2 in murine and chicken  $\delta$ - and  $\gamma$ -crystallin genes (28) and in human *tgf4* (29) has led to recognition of possible protein-protein interactions between Sox2 and adjacent DNA-bound factors, including the Oct-3/4 POU domain. Analysis of the murine *tgf-4* enhancer in an embryonal carcinoma cell line has demonstrated that synergistic activation of the enhancer by Sox2 and Oct-3/4 requires a specific arrangement of factor-binding sites (30). Synergy is mediated by at least two mechanisms, 1) cooperative DNA binding by the HMG box and POU domain (30) and 2) reciprocal conformational changes extending to regions outside of the respective DNA-binding domains and leading to enhanced transcriptional activation function (31). Analogous mechanisms may underlie functional synergy between Sox10 and the classical Zn finger protein Sp1 in transcriptional activation of genes encoding subunits of the neuronal nicotinic acetylcholine receptor (32). The nature of these interactions has not been defined. Individual Sox families often exhibit conservation outside of the HMG box (*i.e.* within other recognizable sequence motifs such as a leucine zipper or serine-threonine-rich region). These regions include classical domains of transcriptional activation or repression; therefore, Sox proteins can, in principle, function as spe-

cific transcription factors (14, 15). Functional evidence in one case is provided by the association between campomelic dysplasia and truncation of or mutations within SOX9's transactivation domain (18, 19). In contrast, almost all mammalian Sry proteins lack discrete transactivation or repression domains. Although sequences outside of the Sry HMG box are generally divergent (22, 33), its extreme C terminus contains a PDZ-binding sequence. A candidate SRY-interacting PDZ protein (SRY interacting protein 1; SIP1) has been identified by the yeast two-hybrid assay (34). Involvement of C-terminal sequences in SRY-mediated gene regulation would rationalize the case report of 46,XY sex reversal associated with an SRY mutation causing deletion of the C-terminal 41 residues but sparing the HMG box (35). Target genes for human SRY or other mammalian Sry proteins are not presently known.

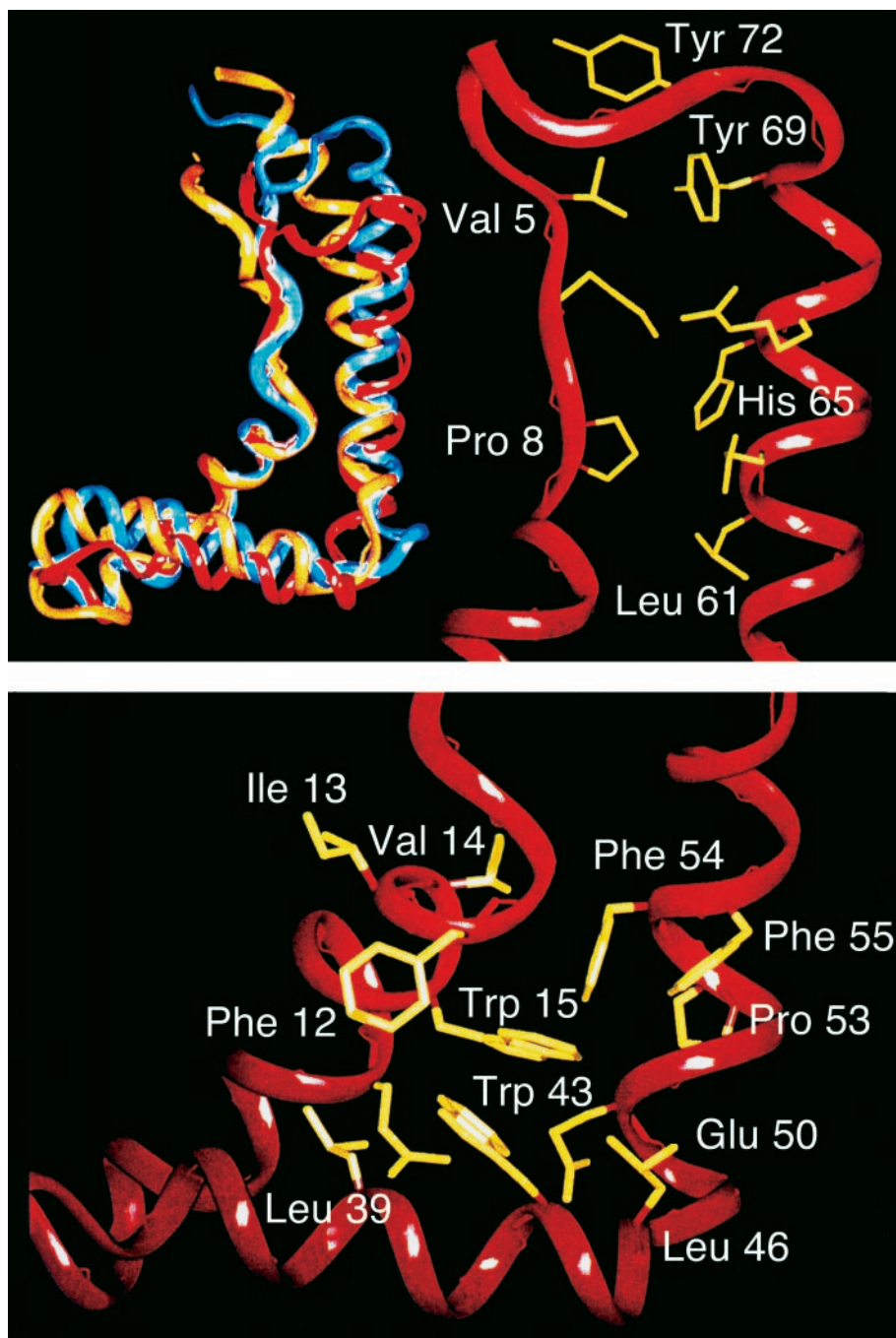
### STRUCTURE OF THE NONSPECIFIC HMG BOX AND DNA COMPLEXES

The purpose of this mini-review is to highlight the unusual dynamics of the Sox HMG box (36, 37) and its possible implications for function. An important foundation is provided by extensive nuclear magnetic resonance (NMR) and crystallographic analyses of nonspecific HMG boxes. Solution structures of four free nonspecific HMG boxes (the A and B domains of HMG-1, *Drosophila* chromosomal protein HMG-D, and yeast protein HNP6A) have been determined (3–6). Structures of these domains are well defined in the absence of DNA. Unlike conventional globular domains, the two wings of the HMG box contain discrete hydrophobic cores. The primary core, located between helix 1 and helix 2, stabilizes the confluence of the major wing. Its organization is dominated by conserved aromatic-aromatic interactions as illustrated in the *bottom panel* of Fig. 2. Such interactions not only contribute to the hydrophobic character of the core but may also impose geometric constraints on packing. Possible packing arrangements of side chains is constrained by the size and planarity of the aromatic rings and by weakly polar electrostatic interactions. The latter involve the  $\pi$ -electrons circulating above and below the faces of the aromatic rings and the partial positive charges of C-H groups at ring edges. A second and less extensive mini-core occurs in the minor wing between  $\alpha$ -helix 3 and the N-terminal  $\beta$ -strand (*upper panel* of Fig. 2). Both wings contribute to the motif's angular DNA-binding surface (7–11). HMG-D (4, 10) and HNP6A (11) are similar to the B domain of HMG1 (5, 6) whereas the A domain (3) differs in the structure of helix 1 and length of the loop between helix 1 and 2. The functional implications of these differences are not well understood. Investigation of main-chain dynamics by analysis of  $^{15}\text{N}$  heteronuclear relaxation times suggests that the nonspecific HMG box can be appropriately described as a rigid, axially symmetric ellipsoid (38).

Crystal structures have been obtained for two nonspecific complexes, 1) the A domain of HMG1 with a 20-bp DNA site containing a *cis*-platin adduct (9); and 2) HMG-D bound to an unmodified decamer DNA site (10). In addition, an NMR-derived model of a nonspecific HNP6A-DNA complex has also been described (11). The structures of bound and free nonspecific HMG boxes are similar. Small structural adjustments occur, presumably to accommodate the details of the distorted DNA surface. The protein-DNA interface is remarkable for nonpolar contacts between the motif's angular protein surface and the DNA's expanded, underwound, and bent minor groove. A conserved wedge of aliphatic and aromatic side chains inserts between successive base pairs to disrupt base stacking as observed in SRY (39, 40). Such partial intercalation is similar to that observed in DNA complexes of the TATA-binding protein (TBP) (41, 42), LacI/PurR family of repressors (43, 44), and hyperthermophilic archaeal chromosomal proteins (45). Although these proteins are unrelated in overall structure, use of a cantilever side chain provides a common mechanism by which base stacking is disrupted at a DNA kink. This side chain inserts *between* base pairs, unlike conventional contacts between side chains and *edges* of bases. At the site of insertion base pairing is, in general, maintained. Of HMG boxes, the structure of the HMG-D complex (10) is particularly remarkable for its three distinct sites of partial intercalation, yielding an overall bend angle of  $111^\circ$ . The interface also contains three water-mediated hydrogen bonds between bases and polar or charged side chains. Water is proposed to function as an adaptor between a given protein side chain and variable target bases, making possible alternate interactions by a common DNA-binding surface. Together, multiple and distributed protein-DNA contacts in the HMG-D complex give rise to a smooth overall DNA bend. A contrasting mode of binding is observed in the complex between HMG-1A and the *cis*-platin-DNA adduct (9). Binding and bending occur predominantly at the site of chemical modification, which induces a preexisting kink in the DNA. The structural basis of recognition of distorted DNA structures by nonspecific HMG domains has recently been reviewed (46, 47).

### STRUCTURES OF SPECIFIC COMPLEXES

The structures of complexes between the specific HMG boxes of SRY and Lef-1 (a non-Sox-specific domain with related specificity 5'-TTCAAA-3'; non-consensus SRY nucleotides in *bold*) and their cognate DNA sites have been determined by NMR spectroscopy (Fig. 3, A and B; Refs. 7 and 8). The structures of the bound HMG boxes strongly resemble those of nonspecific HMG boxes. As anticipated by homology, the SRY- and Lef-1 HMG boxes are each L-shaped and contain discrete wing-specific hydrophobic cores

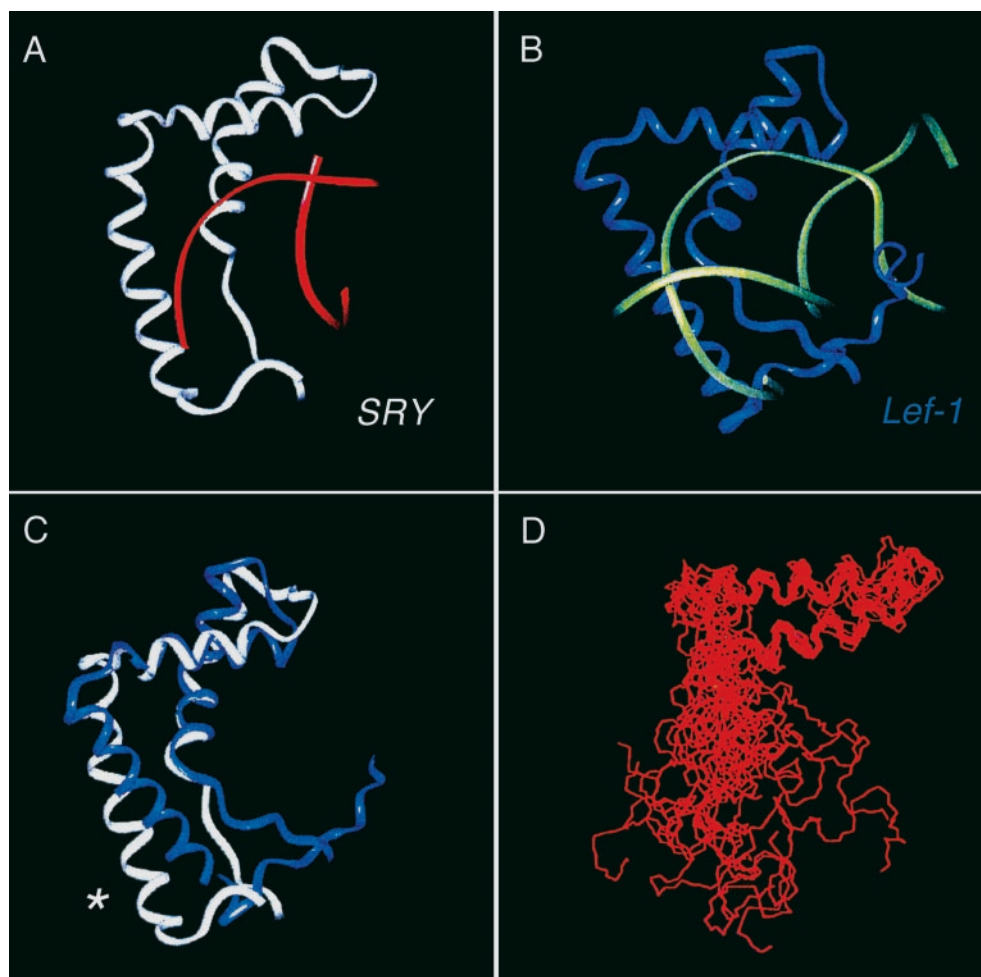


**Fig. 2.** Overview of Side-Chain Packing in the SRY HMG Box

*Upper panel,* A comparison of the bound SRY box (red; Ref. 7) and nonspecific boxes (HMG-D, blue; and HMG-1B, gold) is shown at left. Details of side-chain packing in the minor wing of the bound SRY HMG box are shown at right. *Lower panel,* Structure of the major wing of SRY in a specific complex showing packing of multiple aromatic rings and aliphatic side chains. The details of such packing differ in Lef-1 (8) and nonspecific HMG boxes (3–6).

(Fig. 2, A and B). Comparison of specific and nonspecific complexes has provided insight into the origins of sequence specificity and differences in the position or extent of partial side-chain intercalation (10). Remarkably, such properties seem to reflect sequence changes at only a handful of protein positions. Two examples illustrate a common theme: 1) Lef-1, SRY, and Sox domains contain an invariant Asn at position

10. The Asn carboxamide makes sequence-specific bidentate hydrogen bonds to edges of base pairs at an invariant 5'-TG-3' step in target DNA sites (7, 8). The corresponding side chain in nonspecific domains is Ser10, which also contacts DNA but without sequence specificity. Its interactions in the HMG-D complex, described as sequence neutral, are water-mediated (10). 2) Residues 32, 33, and 36 are nonpolar and



**Fig. 3.** Comparison of Sequence-Specific HMG Boxes

A, Ribbon model of specific SRY-DNA complex (7). The protein is shown in *white* and DNA backbone in *red*. B, Ribbon model of specific Lef1-DNA complex (8). The protein is shown in *blue* and DNA backbone in *green*. C, Superposition of SRY and Lef1 HMG boxes according to the main-chain atoms of  $\alpha$ -helices 1 and 2 demonstrates relative displacement of the minor wing (*asterisk*). D, NMR-derived ensemble of Sox-4 contains well ordered major wing and disordered minor wing (37). Although the N-terminal segment is locally disordered, helix 3 is locally ordered but lacks a coherent orientation relative to the major wing. Details of packing between helix 1 and helix 2 apparently differ from those of other HMG boxes.

likewise sequence neutral in nonspecific domains but polar and capable of specific contacts in specific domains. An example is provided by Phe32 in HMG-1A, which inserts into the *cis*-platin-induced DNA kink. Similarly, Val32 partially intercalates in the HMG-D complex. The corresponding side chain in SRY and Lef-1 is serine, the polarity of which apparently precludes partial intercalation. In the future, the proposed relationship between protein sequence and sequence specificity (10, 46) can be tested by mutagenesis.

Specific SRY and Lef-1 complexes exhibit overall similarities as well as key apparent differences. Each exhibits a single side-chain cantilever at corresponding positions: partial intercalation by Ile (SRY; position 13 of the HMG box consensus, *lower panel* of Fig. 2) or Met (Lef-1) similarly disrupts base stacking but not base pairing (7, 8, 39, 40). Additional sites of insertion as defined in nonspecific complexes (9–11) are not observed. The reported orientation of aromatic side

chains in the major core of SRY differs in detail from that of Lef-1, which is similar to nonspecific HMG boxes. Specific SRY and Lef-1 complexes also differ in apparent bend angle. The 15-bp DNA duplex employed in the Lef-1 complex is bent by approximately  $110^\circ$  and exhibits a remarkable similarity to the corresponding portion of the nonspecific HMG-D complex. The 8-bp DNA duplex employed in the SRY complex is less bent ( $40^\circ$ – $80^\circ$ ); however, its limited length inhibits accurate assessment of the bend angle (J. Love and P. E. Wright, personal communication). The marked difference in extent of DNA bending is in part due to the influence of Lef-1's basic tail, which binds across the major groove as an electrostatic clamp (8). The different DNA bend angles are associated with a change in the orientation between major and minor wings as illustrated by molecular modeling. Superposition of  $\alpha$ -helices 1 and 2 in SRY and Lef-1 gives rise to a large relative displacement in the apparent posi-

tion of  $\alpha$ -helix 3 (*asterisk* in Fig. 3C). Because structures of free SRY and Lef-1 have not independently been determined, it is not known whether the apparent differences between their bound structures result from differential induced fit or instead preexist in the respective unbound proteins.

## STRUCTURE AND DYNAMICS OF SOX DOMAINS

The solution structures of lymphocyte transcriptional activator Sox-4 and testis-specific factor Sox5 in the absence of DNA have been found to exhibit a novel combination of order and disorder (Fig. 3D). The three canonical  $\alpha$ -helices of the HMG box are present and locally well ordered (36, 37). Whereas the tertiary structure of the major wing is well defined in Sox4, the minor wing is not. Similar features occur in free SRY (E. Rivera, N. Phillips, and M. A. Weiss, unpublished results). The major wing's characteristic aromatic-aromatic interactions are associated with dispersion of NMR chemical shifts (the inequivalence of precise proton resonance frequencies due to differences in local environments in a protein) and short-range distances between neighboring side chains in space (nuclear Overhauser enhancements; NOEs). These NMR features are similar in free domains and in specific DNA complexes. The minor wing's characteristic chemical shifts and NOEs are, by contrast, absent in spectra of the free domains. These include otherwise prominent interactions among the side chains of Val5, His65, Tyr69, and Tyr72 (*upper panel* of Fig. 2), residues conserved among Sox sequences (Fig. 4). An illustrative example is provided by ring currents generated by aromatic rings in  $\alpha$ -helix 3 of SRY (Fig. 5). Ring currents, local magnet fields arising from aromatic electrons, are readily estimated by a parameterized dipole approximation (48). In the bound state, such ring currents intersect with the N-terminal  $\beta$ -strand due to folding of the minor wing. In particular, the  $\gamma$ -methyl groups of Val5 overlay the aromatic ring of Tyr69,

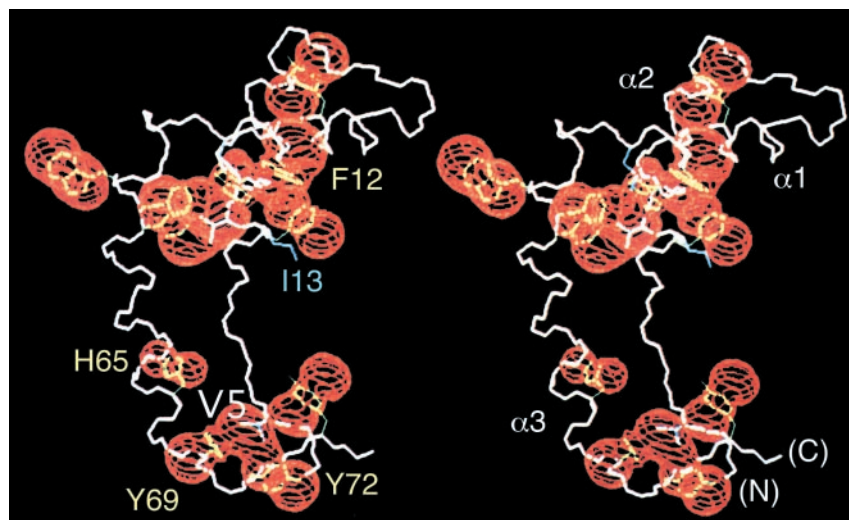
giving rise to a large up-field ring-current shift (*bold-face* values in Table 2) associated with long-range NOEs. None of these features are observed in spectra of the free SRY or Sox4 domains: instead minor-wing side chains, such as Val5 and Tyr72, exhibit motional narrowing, near-random coil chemical shifts, and an absence of long-range NOEs. Because chemical shifts of minor-wing aromatic and methyl resonances can readily be obtained even in the absence of exhaustive NMR analysis, we suggest that these features will be of general value in screening other Sox domains for minor-wing disorder and induced fit on DNA binding.

An ensemble of NMR-based models of Sox4 (37), obtained by distance geometry and simulated annealing (DG/SA), in fact contains no fixed relationship between  $\alpha$ -helix 3 and the major wing ( $\alpha$ -helices 1 and 2). The N-terminal strand is disordered and detached from  $\alpha$ -helix 3. The major hydrophobic core with its conserved aromatic side chains is well organized whereas the minor hydrophobic core is absent. Packing of the N-terminal strand of Sox4 against  $\alpha$ -helix 3 is induced on specific DNA binding (E. Rivera, N. Phillips, and M. A. Weiss, unpublished results). The presence of ordered  $\alpha$ -helical segments with imprecise tertiary relationship is reminiscent of a molten globule (49), an intermediate state of protein organization observed in protein-folding pathways. A schematic model of an equilibrium between open and closed minor wings is provided in Fig. 1B. We caution that the precision of DG/SA models reflects the number of restraints and may or may not correspond to physical fluctuations. The minor wing's imprecision as seen in the Sox4 model (Fig. 3D) thus reflects a paucity of NMR-derived restraints in this region of the protein. The short range and steep distance dependence of the NOE (typically  $<5 \text{ \AA}$  and scaling with  $r^{-6}$ ) implies that distances longer than this cut off cannot routinely be measured. Thus, spatial separations of  $8 \text{ \AA}$  may appear similar to spatial separations of  $20 \text{ \AA}$  as each would give rise to an unobserved signal. Because absence of evidence does not necessarily imply evidence of absence, the DG/SA calculation is underdetermined and hence

	5	*	65	69	72
<i>hSRY</i>	QDR <b>V</b> KRPMNAF I VWS		QKLQAM <b>H</b> REK <b>Y</b> PN <b>Y</b> KYRPRRKAKM		
<i>mSry</i>	EGH <b>V</b> KRPMNAF MVWS		QRLK I L <b>H</b> REK <b>Y</b> PN <b>Y</b> KYQPHRRRAKV		
<i>mSox1</i>	QDR <b>V</b> KRPMNAF MVWS		KRLRAL <b>H</b> MKE <b>H</b> PD <b>Y</b> KYRPRRKTKT		
<i>mSox2</i>	PDR <b>V</b> KRPMNAF MVWS		KRLRAL <b>H</b> MKE <b>H</b> PD <b>Y</b> KYRPRRKTKT		
<i>mSox3</i>	QDR <b>V</b> KRPMNAF MVWS		KRLRAV <b>H</b> MKE <b>Y</b> PD <b>Y</b> KYRPRRKTKT		
<i>mSox4</i>	SGH I KRPMNAF MVWS		ERLRLK <b>H</b> MAD <b>Y</b> PD <b>Y</b> KYRPRKKUKS		
<i>mSox5</i>	EPH I KRPMNAF MVWA		GSLSKQ <b>H</b> L EK <b>Y</b> PD <b>Y</b> KYKPRPKRTC		
<i>mSox6</i>	EPH I KRPMNAF MVWA		ARLSKI <b>H</b> L EK <b>Y</b> PN <b>Y</b> KYKPRPKRTC		
<i>hSox9</i>	KPH <b>V</b> KRPMNAF MVWA		ERLRVQ <b>H</b> KKD <b>H</b> PD <b>Y</b> KYQPRRRKSV		
	$\beta$ strand		$\alpha$ -helix 3		basic tail

**Fig. 4.** Comparison of N-Terminal and C-Terminal Sequences of Human and Murine SRY and Selected Sox Domains

The conservation of Val or Ile at position 5 and an aromatic side chain at position 69 is highlighted (*boxes*). These side chains provide valuable NMR markers for the dynamics and folding of the minor wing (see *upper panel* of Fig. 2). *Asterisk* indicates position 13 of HMG-box consensus, which residue inserts between base pairs as a cantilever to disrupt base stacking (7, 36, 39, 40).



**Fig. 5.** Simulation of Aromatic Ring Currents in the Bound Structure of the SRY HMG Box

Stereo representation of the protein backbone (*white*) and selected side chains. *Red balls* represent contours at an up-field ring current of 0.5 ppm; negative ring currents are not displayed. At bottom a cluster of four aromatic ring currents from C-terminal residues (His65, Tyr69, Tyr72, and Tyr74) is seen to impinge upon the neighboring N-terminal  $\beta$ -strand. The side chain of Val5 is shown in *white* encased within the ring current of Tyr69 (see Table 2).

unphysical. The model shown in Fig. 3D was thus proposed as a working hypothesis rather than definitive characterization of the extent of disorder in the minor wing.

Evidence for an equilibrium between open and closed conformations has been obtained in studies of Sox5 by an elegant combination of biophysical techniques. Although NMR studies likewise suggested that its minor wing is largely unfolded at 37 C, decreasing temperature was found to lead to progressive folding of this segment (36). Intensities of key interresidue NOEs, chosen to reflect tertiary contacts, were monitored as a function of temperature. Whereas the intensity of NOEs diagnostic of the major wing's tertiary structure was unaffected by temperature in the range 16–31 C, attenuation of minor wing-specific NOEs was observed with increasing temperature. Although these NMR observations in themselves could have multiple interpretations, complementary evidence of discrete major and minor wing unfolding transitions was obtained by fluorescence spectroscopy and differential scanning calorimetry. The minor wing of Sox5 unfolds with a midpoint of 34 C whereas the major wing unfolds with a midpoint of 46 C (36). Analogous studies of the free SRY HMG box suggests that its minor wing is unfolded even at temperatures as low as 4 C (N. Phillips and M. A. Weiss, unpublished results). Unfortunately, none of these methods can provide a quantitative estimate of the extent of unfolding. Although the unusual spectroscopic features of Sox domains emphasize the distinction between the dynamics of the two wings, NMR and fluorescent studies have not to date addressed the extent of excursions between the domain's N-terminal segment and helix 3, *i.e.* how open is the "open" state? In particular, because the NMR methods employed in these studies are based

on the short-range NOE interaction (48), it is possible, in principle, that more long-range order is present in solution than is suggested by the DG/SA model shown in Fig. 3D. In the future it would be of interest to investigate the extent of long-range correlation between the major wing and  $\alpha$ -helix 3 by use of residual dipolar couplings in partially oriented samples in solution. This new NMR methodology (50) circumvents the restriction of NMR parameters to local properties or short-range interactions (48). It would be of complementary interest to measure distributions of long-range distances by time-resolved fluorescence resonance energy transfer (FRET) (51, 52). Although FRET is formally a  $r^{-6}$  dipole-dipole interaction like the NOE, its distance range is determined by the Förster distance ( $R_0$ ) governing resonance energy transfer between donor and acceptor probes. This distance is probe dependent and typically lies in the range 10–80 Å. Attachment of suitable probes to the N-terminal segment and helix 3 would thus enable direct characterization of long-range distances and fluctuations. It is likely that time-resolved FRET analysis of Sox5 would permit a definitive test of the hypothesis that the free domain exists in an equilibrium between open and closed conformations.

The DNA-dependent order-disorder transition of Sox domains differs in kind from those of basic zipper (bZIP) and basic helix-loop-helix (bHLH) major-groove DNA-binding motifs (53–57). The latter contain disordered N-terminal basic arms, which form divergent pairs of recognition  $\alpha$ -helices on specific DNA binding. Induced fit in the protein thus occurs at the level of secondary structure. Although small conformational adjustments occur in DNA structure, including limited DNA bending (typically  $<20^\circ$ ), the DNA remains in the B family, and its major groove acts essentially as a

**Table 2.** Influence on V5 of Predicted Aromatic Ring Currents in SRY Complex<sup>a</sup>

Proton	F12	W15	Y41	W43	W52	F54	F55	H65	Y69	Y72	Y74	Sum
V5 H <sub>α</sub>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	<b>0.27</b>	0.12	0.05	<b>0.46</b>
V5 H <sub>β</sub>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	<b>0.36</b>	0.01	0.16	<b>0.53</b>
V5 <sub>γ1</sub> CH <sub>3</sub>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	<b>1.67</b>	0.10	0.06	<b>1.84</b>
V5 <sub>γ2</sub> CH <sub>3</sub>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	<b>0.36</b>	<b>0.29</b>	0.14	<b>0.78</b>

<sup>a</sup> Values shown represent the average of individual calculations for each member of the published ensemble. For methyl resonances the value is further averaged among the three component protons. Influence of the Y69 ring current (helix 3) on V5 (N-terminal strand), reflecting long-range contact (Fig. 1C), is shown in *boldface*; a smaller influence is predicted due to the Y72 ring current.

**Table 3.** DNA Bending by Sequence-Specific HMG Boxes

DNA Site <sup>a</sup>	HMG Box	Angle <sup>b</sup>	DNA Site <sup>a</sup>	HMG Box	Angle <sup>b</sup>
Site 1	hSRY	30°	Site 1	LEF-1	85°
Site 2	hSRY	60°	Site 2	LEF-1	85°
Site 1	mSry	60°	Site 3	LEF-1	90°
Site 2	mSry	85°	Site 4	LEF-1	130°

<sup>a</sup> The following DNA sequences were employed in permutation gel electrophoresis (60, 61): site 1, 5'-CCCATTGTTGCT-3' and complement (designated TCR-15 in Refs. 60 and 61); site 2, 5'-CCCATTGTTCTC-3' and complement (designated TCR-11); site 3, 5'-CCCTTTGTTTCT-3' and complement (designated TCR-8); and site 4, 5'-CCCTTTGAAGCT-3' and complement (designated TCR-1).

<sup>b</sup> DNA bend angles were estimated by permutation gel electrophoresis and have not been independently corroborated by spectroscopic or structural studies.

performed template for protein folding (58, 59). The  $\alpha$ -helical structure of Sox domains is by contrast pre-organized: it is *tertiary* structure that is specified on DNA binding (36, 37, 58). Sox-induced DNA bending thus reflects a *bidirectional* induced fit wherein the Sox domain instructs the DNA how to bend as the DNA instructs the protein to complete its tertiary fold. Because the angular surface of the Sox HMG box is not fixed, a given domain may be compatible with a range of DNA bend angles rather than a single value. This hypothesis suggests that the precise DNA bend adopted in a specific complex could depend on the exact DNA target sequence and presence of neighboring protein-DNA complexes. Indeed, sequence-dependent DNA bending has been inferred from electrophoretic measurements of sequence-specific HMG boxes, including human and murine Sry and LEF1 (Table 3; Refs. 60 and 61). Induced bend angles can differ by as much as 30°, implying a substantial difference in underlying DNA and protein structures. Accordingly, it would be of future interest to obtain crystal or NMR structures of the same Sox HMG box bound to variant DNA sites associated with different electrophoretic bend angles. Analogous crystallographic studies of the TBP in variant DNA complexes revealed no changes in bend angle, suggesting that TBP (unlike a Sox domain) functions as a robust and preformed template for DNA bending (62).

Why are specific HMG boxes floppy? We propose that adaptability of the motif's angular surface enables a Sox protein to induce different architectures in different functional contexts. We imagine that target genes for a given factor will differ, for example, in the precise sequence of Sox binding sites and its combinatorial relation to other factor-binding sites in the same promoter or enhancer. Because context-dependent changes in overall architecture may differentially affect transcription, a single factor may exert fine con-

trol over relative levels of expression within a set of target genes. Were the specific HMG box a rigid platform directing a preset DNA bend angle, then the cell might need a very large collection of factors, each calibrated to a different angle, to effect such fine control. Use of a single flexible motif with an adjustable set point for DNA bending would represent a striking economy of protein design. Testing this hypothesis will require the *in vitro* reconstitution and structural characterization of Sox-specific enhanceosomes.

### Acknowledgments

We thank A. S. Stern for ring-current shift calculations; D. N. Jones and N. Narayana for assistance with figures; G. Chen, E. Haas, A. Jancso, J. Radek, E. Rivera, and N. Phillips for communication of results before publication; and P. Donahoe, C.-Y. King, J. Love, and P. E. Wright for discussion.

Received October 13, 2000. Revision received December 1, 2000. Accepted December 20, 2000.

Address requests for reprints to: Dr. Michael A. Weiss, Department of Biochemistry, Case Western Reserve University School of Medicine, 2109 Adelbert Road, Cleveland, Ohio 44106-4935.

This work was supported in part by NIH Grant CA-63485 to M.A.W.

**Note Added in Proof.** Recent studies of the murine Sox-5 HMG box by multidimensional NMR and <sup>15</sup>N relaxation measurements have demonstrated anomalous mobility of the minor wing (63).

### REFERENCES

1. Ner S 1992 HMGs everywhere. *Curr Biol* 2:208–210
2. Bewley CA, Gronenborn AM, Clore GM 1998 Minor groove-binding architectural proteins: structure, function, and DNA recognition. *Annu Rev Biophys Biomol Struct* 27:105–131



3. Hardman CH, Broadhurst RW, Raine AR, Grasser KD, Thomas JO, Laue ED 1995 Structure of the A-domain of HMG1 and its interaction with DNA as studied by heteronuclear three- and four-dimensional NMR spectroscopy. *Biochemistry* 34:16596–16607
4. Jones DN, Searles MA, Shaw GL, Churchill ME, Ner SS, Keeler J, Travers AA, Neuhaus D 1994 The solution structure and dynamics of the DNA-binding domain of HMG-D from *Drosophila melanogaster*. *Structure* 2:609–627
5. Read CM, Cary PD, Crane-Robinson C, Driscoll PC, Norman DG 1993 Solution structure of a DNA-binding domain from HMG1. *Nucleic Acids Res* 21:3427–3436
6. Weir HM, Kraulis PJ, Hill CS, Raine AR, Laue ED, Thomas JO 1993 Structure of the HMG box motif in the B-domain of HMG1. *EMBO J* 12:1311–1319
7. Werner MH, Huth JR, Gronenborn AM, Clore GM 1995 Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. *Cell* 81:705–714
8. Love JJ, Li X, Case DA, Giese K, Grosschedl R, Wright PE 1995 Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature* 376:791–795
9. Ohndorf UM, Rould MA, He Q, Pabo CO, Lippard SJ 1999 Basis for recognition of cisplatin-modified DNA by high-mobility-group proteins. *Nature* 399:708–712
10. Murphy, F.V.t., Sweet RM, Churchill ME 1999 The structure of a chromosomal high mobility group protein-DNA complex reveals sequence-neutral mechanisms important for non-sequence-specific DNA recognition. *EMBO J* 18:6610–6618
11. Allain FH, Yen YM, Masse JE, Schultze P, Dieckmann T, Johnson RC, Feigon J 1999 Solution structure of the HMG protein NHP6A and its interaction with DNA reveals the structural determinants for non-sequence-specific binding. *EMBO J* 18:2563–2579
12. Grosschedl R 1995 Higher-order nucleoprotein complexes in transcription: analogies with site-specific recombination. *Curr Opin Cell Biol* 7:362–370
13. Kornberg RD, Lorch Y 1995 Interplay between chromatin structure and transcription. *Curr Opin Cell Biol* 7:371–375
14. Pevny LH, Lovell-Badge R 1997 Sox genes find their feet. *Curr Opin Genet Dev* 7:338–344
15. Wegner M 1999 From head to toes: the multiple facets of Sox proteins. *Nucleic Acids Res* 27:1409–1420
16. Gubbay J, Collignon J, Koopman P, Capel B, Economou A, Munsterberg A, Vivian N, Goodfellow P, Lovell-Badge R 1990 A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature* 346:245–250
17. Schafer AJ, Goodfellow PN 1996 Sex determination in humans. *Bioessays* 18:955–963
18. Foster JW, Dominguez-Steglich MA, Guioli S, Kowk G, Weller PA, Stevanovic M, Weissenbach J, Mansour S, Young ID, Goodfellow PN, Brook JD, Schafer AJ 1994 Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. *Nature* 372:525–530
19. Wagner T, Wirth J, Meyer J, Zabel B, Held M, Zimmer J, Pasantes J, Briccarelli FD, Keutel J, Hustert E, Wolf U, Tommerup N, Schempp W, Scherer G 1994 Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9. *Cell* 79:1111–1120
20. Schilham MW, Oosterwegel MA, Moerer P, Ya J, de Boer PA, van de Wetering M, Verbeek S, Lamers WH, Kruisbeek AM, Cumano A, Clevers H 1996 Defects in cardiac outflow tract formation and pro-B-lymphocyte expansion in mice lacking Sox-4. *Nature* 380:711–714
21. Russell SR, Sanchez-Soriano N, Wright CR, Ashburner M 1996 The Dichaete gene of *Drosophila melanogaster* encodes a SOX-domain protein required for embryonic segmentation. *Development* 122:3669–3676
22. Koopman P 1999 Sry and Sox9: mammalian testis-determining genes. *Cell Mol Life Sci* 55:839–856
23. Harley VR, Lovell-Badge R, Goodfellow PN 1994 Definition of a consensus DNA binding site for SRY. *Nucleic Acids Res* 22:1500–1501
24. Denny P, Swift S, Connor F, Ashworth A 1992 An SRY-related gene expressed during spermatogenesis in the mouse encodes a sequence-specific DNA-binding protein. *EMBO J* 11:3705–3712
25. Kanai Y, Kanai-Azuma M, Noce T, Saido TC, Shiroishi T, Hayashi Y, Yazaki K 1996 Identification of two Sox17 messenger RNA isoforms, with and without the high mobility group box region, and their differential expression in mouse spermatogenesis. *J Cell Biol* 133:667–681
26. Mertin S, McDowall SG, Harley VR 1999 The DNA-binding specificity of SOX9 and other SOX proteins. *Nucleic Acids Res* 27:1359–1364
27. Peirano RI, Wegner M 2000 The glial transcription factor Sox10 binds to DNA both as monomer and dimer with different functional consequences. *Nucleic Acids Res* 28:3047–3055
28. Kamachi Y, Sockanathan S, Liu Q, Breitman M, Lovell-Badge R, Kondoh H 1995 Involvement of SOX proteins in lens-specific activation of crystallin genes. *EMBO J* 14:3510–3519
29. Yuan H, Corbi N, Basilico C, Dailey L 1995 Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes Dev* 9:2635–2645
30. Ambrosetti DC, Basilico C, Dailey L 1997 Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. *Mol Cell Biol* 17:6321–6329
31. Ambrosetti DC, Scholer HR, Dailey L, Basilico C 2000 Modulation of the activity of multiple transcriptional activation domains by the DNA binding domains mediates the synergistic action of Sox2 and Oct-3 on the fibroblast growth factor-4 enhancer. *J Biol Chem* 275:23387–23397
32. Melnikova IN, Lin H, Blanchette AR, Gardner PD 2000 Synergistic transcriptional activation by Sox10 and Sp1 family members. *Neuropharmacology* 39:2615–2623
33. Whitfield LS, Lovell-Badge R, Goodfellow PN 1993 Rapid sequence evolution of the mammalian sex-determining gene SRY. *Nature* 364:713–715
34. Poulat F, Barbara PS, Desclozeaux M, Soullier S, Moniot B, Bonneaud N, Boizet B, Berta P 1997 The human testis determining factor SRY binds a nuclear factor containing PDZ protein interaction domains. *J Biol Chem* 272:7167–7172
35. Tajima T, Nakae J, Shinohara N, Fujieda K 1994 A novel mutation localized in the 3' non-HMG box region of the SRY gene in 46,XY gonadal dysgenesis. *Hum Mol Genet* 3:1187–1189
36. Crane-Robinson C, Read CM, Cary PD, Driscoll PC, Dragan AI, Privalov PL 1998 The energetics of HMG box interactions with DNA. Thermodynamic description of the box from mouse Sox-5. *J Mol Biol* 281:705–717
37. van Houte LP, Chuprina VP, van der Wetering M, Boelens R, Kaptein R, Clevers H 1995 Solution structure of the sequence-specific HMG box of the lymphocyte transcriptional activator Sox-4. *J Biol Chem* 270:30516–30524
38. Broadhurst RW, Hardman CH, Thomas JO, Laue ED 1995 Backbone dynamics of the A-domain of HMG1 as studied by <sup>15</sup>N NMR spectroscopy. *Biochemistry* 34:16608–16617
39. Haqq CM, King CY, Ukiyama E, Falsafi S, Haqq TN, Donahoe PK, Weiss MA 1994 Molecular basis of mammalian sexual determination: activation of Mullerian inhibiting substance gene expression by SRY. *Science* 266:1494–1500
40. King CY, Weiss MA 1993 The SRY high-mobility-group box recognizes DNA by partial intercalation in the minor groove: a topological mechanism of sequence specificity. *Proc Natl Acad Sci USA* 90:11990–11994

41. Kim JL, Nikolov DB, Burley SK 1993 Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* 365:520–527
42. Kim Y, Geiger JH, Hahn S, Sigler PB 1993 Crystal structure of a yeast TBP/TATA-box complex. *Nature* 365:512–520
43. Lewis M, Chang G, Horton NC, Kercher MA, Pace HC, Schumacher MA, Brennan RG, Lu P 1996 Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* 271:1247–1254
44. Schumacher MA, Choi KY, Zalkin H, Brennan RG 1994 Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by  $\alpha$  helices. *Science* 266:763–770
45. Robinson H, Gao YG, McCrary BS, Edmondson SP, Shriver JW, Wang AH 1998 The hyperthermophile chromosomal protein Sac7d sharply kinks DNA. *Nature* 392:202–205
46. Murphy IV FV, Churchill ME 2000 Nonsequence-specific DNA recognition: a structural perspective. *Structure Fold Des* 8:R83–89
47. Travers A 2000 Recognition of distorted DNA structures by HMG domains. *Curr Opin Struct Biol* 10:102–109
48. Wuthrich K 1986 *NMR of Proteins and Nucleic Acids*. John Wiley & Sons, New York
49. Redfield C, Smith RA, Dobson CM 1994 Structural characterization of a highly-ordered 'molten globule' at low pH. *Nat Struct Biol* 1:23–29
50. Tjandra N, Bax A 1997 Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* 278:1111–1114
51. Millar DP 2000 Time-resolved fluorescence methods for analysis of DNA-protein interactions. *Methods Enzymol* 323:442–459
52. Ratner V, Sinev M, Haas E 2000 Determination of intramolecular distance distribution during protein folding on the millisecond timescale. *J Mol Biol* 299:1363–1371
53. Weiss MA, Ellenberger T, Wobbe CR, Lee JP, Harrison SC, Struhl K 1990 Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature* 347:575–578
54. O'Neil KT, Hoess RH, DeGrado WF 1990 Design of DNA-binding peptides based on the leucine zipper motif. *Science* 249:774–778
55. Patel L, Abate C, Curran T 1990 Altered protein conformation on DNA binding by Fos and Jun. *Nature* 347:572–575
56. Ferre-D'Amare AR, Pognonec P, Roeder RG, Burley SK 1994 Structure and function of the b/HLH/Z domain of USF. *EMBO J* 13:180–9
57. Ellenberger TE, Brandl CJ, Struhl K, Harrison SC 1992 The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted  $\alpha$  helices: crystal structure of the protein-DNA complex. *Cell* 71:1223–1237
58. Benevides JM, Chan G, Lu XJ, Olson WK, Weiss MA, Thomas Jr GJ 2000 Protein-directed DNA structure. I. Raman spectroscopy of a high-mobility-group box with application to human sex reversal. *Biochemistry* 39:537–547
59. Benevides JM, Li T, Lu XJ, Srinivasan AR, Olson WK, Weiss MA, Thomas Jr GJ 2000 Protein-directed DNA structure II. Raman spectroscopy of a leucine zipper bZIP complex. *Biochemistry* 39:548–556
60. Giese K, Pagel J, Grosschedl R 1994 Distinct DNA-binding properties of the high mobility group domain of murine and human SRY sex-determining factors. *Proc Natl Acad Sci USA* 91:3368–3372
61. Giese K, Pagel J, Grosschedl R 1997 Functional analysis of DNA bending and unwinding by the high mobility group domain of LEF-1. *Proc Natl Acad Sci USA* 94:12845–12850
62. Patikoglou GA, Kim JL, Sun L, Yang SH, Kodadek T, Burley SK 1999 TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev* 13:3217–3230
63. Cary PD, Read CM, Davis B, Driscoll PC, Crane-Robinson C 2001 Solution structure and backbone dynamics of the DNA-binding domain of mouse Sox-5. *Protein Sci* 10:83–98