

Structural Basis for SRY-dependent 46-X,Y Sex Reversal: Modulation of DNA Bending by a Naturally Occurring Point Mutation

Elizabeth C. Murphy¹, Victor B. Zhurkin², John M. Louis¹,
Gabriel Cornilescu¹ and G. Marius Clore^{1*}

¹Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases National Institutes of Health Bethesda, MD 20892-0510, USA

²Laboratory of Experimental and Computational Biology Building 12B, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

The HMG-box domain of the human male sex-determining factor SRY, hSRY_{HMG} (comprising residues 57-140 of the full-length sequence), binds DNA sequence-specifically in the minor groove, resulting in substantial DNA bending. The majority of point mutations resulting in 46X,Y sex reversal are located within this domain. One clinical *de novo* mutation, M64I in the full-length hSRY sequence, which corresponds to M9I in the present hSRY_{HMG} construct, acts principally by reducing the extent of DNA bending. To elucidate the structural consequences of the M9I mutation, we have solved the 3D solution structures of wild-type and M9I hSRY_{HMG} complexed to a DNA 14mer by NMR, including the use of residual dipolar couplings to derive long-range orientational information. We show that the average bend angle (derived from an ensemble of 400 simulated annealing structures for each complex) is reduced by ~13° from 54(±2)° in the wild-type complex to 41(±2)° in the M9I complex. The difference in DNA bending can be localized directly to changes in roll and tilt angles in the ApA base-pair step involved in interactions with residue 9 and partial intercalation of Ile13. The larger bend angle in the wild-type complex arises as a direct consequence of steric repulsion of the sugar of the second adenine by the bulky S^δ atom of Met9, whose position is fixed by a hydrogen bond with the guanidino group of Arg17. In the M9I mutant, this hydrogen bond can no longer occur, and the less bulky C^γm methyl group of Ile9 braces the sugar moieties of the two adenine residues, thereby decreasing the roll and tilt angles at the ApA step by ~8° and ~5°, respectively, and resulting in an overall difference in bend angle of ~13° between the two complexes. To our knowledge, this is one of the first examples where the effects of a clinical mutation involving a protein-DNA complex have been visualized at the atomic level.

Keywords: SRY-DNA complex; 46X,Y sex reversal; DNA bending mutation; multidimensional NMR

*Corresponding author

Introduction

Architectural minor groove DNA-binding proteins play a crucial role in modulating DNA bending, thereby regulating the formation of higher-order multiprotein-DNA complexes, such as enhanceosomes and repressosomes, involved in transcription, recombination and replication.^{1–8} It

is the function of these architectural proteins to bring the requisite transcription and ancillary factors bound to distant DNA sites into close proximity by appropriately altering the path of the DNA template. Examples of minor groove DNA-binding architectural proteins include the high mobility group (HMG)-box (or HMGB) family, the HMG-I/Y (or HMGA) family, TATA-binding protein and integration host factor.

The HMG-box domain is approximately 80 residues in length and defines a superfamily of architectural factors that play a central role in mammalian gene regulation and organogenesis.^{4,8–11} The superfamily can be sub-

Abbreviations used: HMG, high mobility group; LEF, lymphoid enhancing factor; NOE, nuclear Overhauser enhancement.

E-mail address of the corresponding author: clore@speck.niddk.nih.gov

divided into sequence-specific DNA-binding proteins, such as the male sex-determining factor SRY, the SOX proteins and lymphoid enhancer factor-1 (LEF), as well as proteins that bind DNA non-specifically such as HMG1/2, HMGD and NHP6A. Structures of complexes of SRY,¹² LEF-1,¹³ HMGD¹⁴ and NHP6A¹⁵ bound to DNA have been solved. In addition, a complex of HMG1 box A bound to cisplatin-modified DNA has been determined.¹⁶ In all cases, binding occurs in the minor groove and induces a large bend ($\sim 60\text{--}100^\circ$) away from the body of the protein in the direction of the major groove.

SRY plays a critical role in male sexual development, acting at the branch-point in mammalian gonadogenesis where the embryonic bipotential gonad proceeds down either the male or female developmental pathway.^{17–20} Naturally occurring mutations in human SRY, the majority of which are located in its HMG-box domain, result in 46X,Y sex reversal; that is, individuals who are genetically male but phenotypically female. Although the direct target genes for SRY *in vivo* have not been determined, consensus DNA-binding sites for SRY have been identified in the promoters for several genes expressed during testicular development, including the *Amh* gene,²¹ whose product anti-Müllerian hormone (otherwise known as Müllerian-inhibiting substance) is responsible for the regression of the female reproductive duct system. Although the *Amh* gene can be activated by SRY in transfected cell lines,²² it appears unlikely that it is a direct target for SRY *in vivo*, since its transcripts appear only after expression of the *Sry* gene has been switched off.²³ At this time, it appears that SRY acts through the intermediary of a number of other genes, possible candidates for which include the genes for SOX9 and steroidogenic factor 1.^{18,20} Indeed, in males the expression of the *Sox9* gene is upregulated immediately following expression of *Sry*, persisting throughout the testis development; in genetic females, on the other hand, where *Sry* is not expressed, expression of *Sox9* is switched off.^{24,25}

Human SRY is a 204 residue nuclear protein comprising three domains: a central HMG-box domain (referred to as hSRY_{HMG}) approximately 80 residues in length, surrounded by N and C-terminal domains. hSRY_{HMG} is highly conserved among mammalian species, while there is much more variation, both in length and sequence, in the N and C-terminal domains.^{18,20,26} In addition to its DNA binding properties, hSRY_{HMG} contains two nuclear localization sites at its N and C termini²⁷ and a calmodulin-binding site that overlaps with the N-terminal nuclear localization site.²⁸

From a structural perspective, point mutations in hSRY_{HMG} that result in 46X,Y sex reversal can be classified into those that affect the stability and structure of the protein and those that alter its DNA-binding properties.^{12,29} Clinically, the mutations can be subdivided into familial inherited mutations with variable penetrance and

de novo mutations with full penetrance. A number of such naturally occurring mutations have been studied.³⁰ In general, familial mutations have only a moderate effect on complex formation, whereas *de novo* mutations tend to be more severe. One such *de novo* mutation, M64I (which corresponds to M9I in the hSRY_{HMG} construct employed here) is of particular interest, since it discriminates between DNA binding and DNA bending.³⁰ Thus, the DNA-binding affinity of the M64I mutant is minimally reduced by, at most, a factor of 3 relative to that of wild-type for a wide range of different target sites; however, the apparent DNA bend angle deduced from circular permutation gel shift assays is approximately 20° less for the M64I complex than the wild-type complex ($\sim 55^\circ$ versus $\sim 75^\circ$).³⁰ These data suggest that the phenotypic effect of the M64I mutation arises predominantly from the reduction in DNA bending. Indeed, a difference of $10\text{--}20^\circ$ in DNA bend angle, while seemingly small, will result in large translational displacements of proteins bound even a few turns of the DNA helix away from the SRY binding site. Consequently, the spatial relationships of proteins within the nucleoprotein complex organized by SRY will be altered dramatically by the M64I mutation, to the extent that the formation of a transcriptionally competent nucleoprotein complex may no longer be possible. Here, we address directly the structural consequences of the M64I mutation by solving the 3D solution structures of wild-type and mutant hSRY_{HMG} complexed to DNA by multidimensional NMR. This represents a particularly challenging task for NMR, since the anticipated structural differences are rather subtle and hence require the generation of precise and accurate structures.

Results and Discussion

Structure determination

The hSRY_{HMG} construct employed in this work spans residues 57–140 of full-length hSRY, which correspond to residues 2–85 in the present numbering scheme, plus an N-terminal methionine residue. We have solved the solution structure of wild-type hSRY_{HMG} and the M9I mutant (which is equivalent to the M64I mutation in full-length hSRY) complexed to a 14mer duplex DNA comprising the sequence 5'd(CCTGCACAAAC ACC).5'd(GGTGTTTGTGCAGG), which contains the specific DNA-binding site (in bold) for SRY in the promoter of the *Amh* (also known as *Mis*) gene.

The original structure of the HMG-box of hSRY complexed to DNA¹² was carried out on a shorter protein construct (residues 57–133 of full-length hSRY) and a shorter piece of DNA (the central 8 bp of the present sequence). The DNA octamer was too short to permit an accurate determination of the bend angle and, moreover, the HMG-box

domain extended beyond one end of the octamer DNA fragment, leaving the possibility of additional protein-DNA interactions. In addition, interactions between the C-terminal tail and the major groove were observed in the related structure of the complex of LEF-1 and DNA¹³ that were not observed in the original SRY-DNA complex,¹² suggesting that the protein domain employed may have been too short. For these reasons, the present work made use of both a longer fragment of the HMG-box domain (extended by seven residues at the C terminus) as well as a longer DNA fragment (14mer *versus* 8mer).

A circular permutation gel-shift assay was used to verify that the magnitude of DNA bending observed with the present DNA target site and hSRY_{HMG} constructs was consistent with that observed previously³⁰ using slightly different DNA sequences and a shorter HMG-box construct (corresponding to residues 1-77 of the present construct). The same results, within experimental error, were obtained as those reported by Pontiggia *et al.*,³⁰ confirming an apparent difference in bend angle of ~20° between the wild-type and M9I hSRY_{HMG} complexes.

The structure of the wild-type and M9I hSRY_{HMG} complexes with the 14mer DNA were solved by multidimensional heteronuclear NMR spectroscopy³¹⁻³³ using both ¹⁵N and ¹³C-labeled protein and oligonucleotide. Both complexes are in slow exchange on the chemical shift time-scale. Experiments were carried out on the following 1:1 complexes: ¹⁵N-hSRY_{HMG} and unlabeled DNA, ¹⁵N/¹³C-hSRY_{HMG} and unlabeled DNA, and ¹⁵N-hSRY_{HMG} and ¹⁵N/¹³C-DNA. An example of the quality of the data is provided by a series of strips taken from 3D ¹³C-separated/¹²C-filtered nuclear Overhauser enhancement (NOE) experiments illustrating intermolecular NOEs (i.e. short interproton distance contacts) involving Met9 in the wild-type complex (Figure 1(a)) and Ile9 in the M9I complex (Figure 1(b)). In addition to structural information that is reliant on atoms in close spatial proximity (i.e. ≤ 6 Å), as exemplified by NOE, three-bond coupling and chemical shift data, the structure determination also made use of residual heteronuclear (¹D_{NH}, ¹D_{CH}, ¹D_{NC} and ²D_{HNC}) and homonuclear (¹D_{HH}) dipolar couplings measured in a dilute liquid crystalline medium of bicelles³⁴ to provide long-range orientational information, thereby increasing the accuracy of the structure determination.³⁵⁻⁴⁰

The wild-type and M9I hSRY_{HMG}-DNA complexes were solved on the basis of 2798 and 2706 experimental NMR restraints, respectively. The restraints included 168 intermolecular NOE-derived interproton distance restraints for each complex, which is over twice that identified in the original Werner *et al.* structure,¹² as well as 375 and 353 residual dipolar coupling restraints for the wild-type and M9I complexes, respectively. A summary of the structural statistics is provided in Table 1, and superpositions of the final ensemble

of 400 structures each for the wild-type and M9I complexes are displayed in Figure 2.

The structure of hSRY_{HMG}

hSRY_{HMG} has an L-type shape (Figure 3) and consists of extended strands at the N (residues 4-10) and C (residues 72-81) termini, and three helices (residues 11-26, 32-44 and 48-68). The N-terminal three residues (1-3) and the C-terminal four residues (82-85) are disordered in solution. The long arm of the L is made up of the N-terminal tail and helix 3, which are oriented anti-parallel to each other, while the short arm is formed by helices 1 and 2. The C-terminal tail is directed at about right-angles to the N-terminal tail and helix 3. The three helices are approximately mutually orthogonal to each other (with inter-helix angles of 123°, 117° and 113° between helices 1 and 2, 1 and 3 and 2 and 3, respectively). The N-terminal tail is anchored to the end of helix 3 and the beginning of the C-terminal tail by a hydrophobic cluster consisting of Val5, His65, Tyr69 and Tyr72 (Figure 3(b)). There are also good packing interactions between the N-terminal strand (Pro8), the N-terminal end of helix 1 (Ala11 and Val14) and the inner face of helix 3 (Phe55, Ala58 and Leu61). The core formed at the intersection of the three helices consists of three mutually orthogonal aromatic rings (Trp15, Trp43 and Phe54) surrounded by Leu39, Leu46, Lys51 and Phe55 (Figure 3(b)). The orientation of helices 1 and 2 is also stabilized by the packing of Met23, Ala24, Met30, Ile35 and Leu39 (Figure 3(b)).

Wild-type and M9I hSRY_{HMG} are identical at the secondary structure level. The overall backbone atomic r.m.s. difference between the two restrained regularized mean structures for residues 4-81 is 0.86 Å. Regions of significant difference, which are principally translational in nature, are apparent at the N terminus around the site of mutation (residues 4-10), in the loop connecting helices 1 and 2 (residues 27-31), and in the C-terminal tail (residues 69-81) (Figures 3(a) and 4(a)). Excluding these three regions, the best-fit backbone atomic r.m.s. difference between the two proteins (residues 11-26 and 32-68) is reduced to 0.33 Å, which is well within the errors of the coordinates (Figure 4(a)), and the resulting backbone atomic r.m.s. displacements for residues 4-10, 27-31 and 69-81 are 0.9, 0.8 and 2.5 Å, respectively. All three regions of difference are characterized by either low density or absent (in the case of residues 74-81) intramolecular contacts with the protein core (which comprises residues 5-72, Figure 3(b)), and hence have the potential to readily accommodate the M9I mutation itself (in the case of the N-terminal tail) or differences in DNA bend angle (in the case of the loop connecting helices 1 and 2 and the C-terminal tail, located at the top and bottom, respectively, of the DNA-binding site), thereby optimizing intermolecular protein-DNA contacts in the two complexes (see Figure 3(a)).

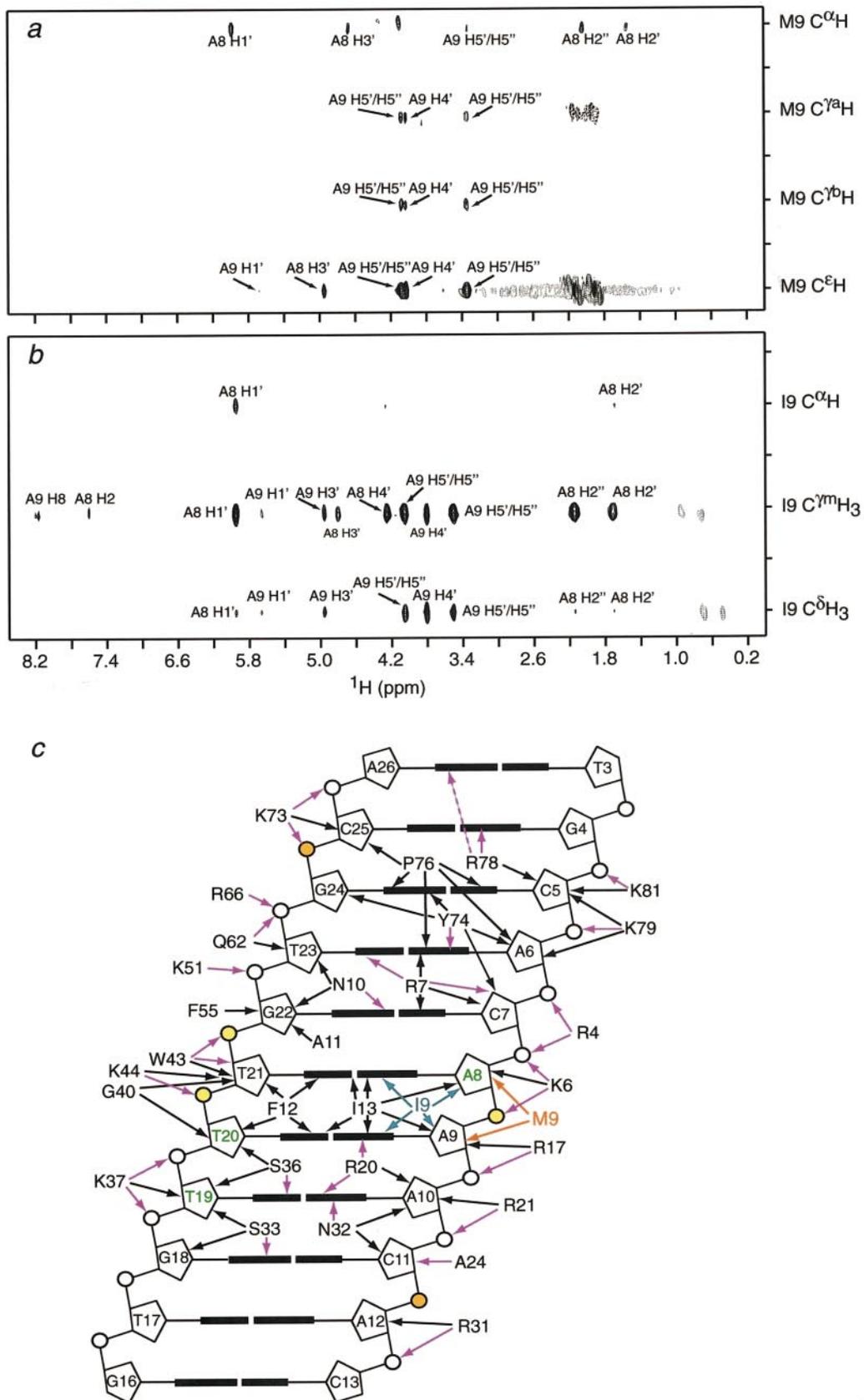


Figure 1 (legend opposite)

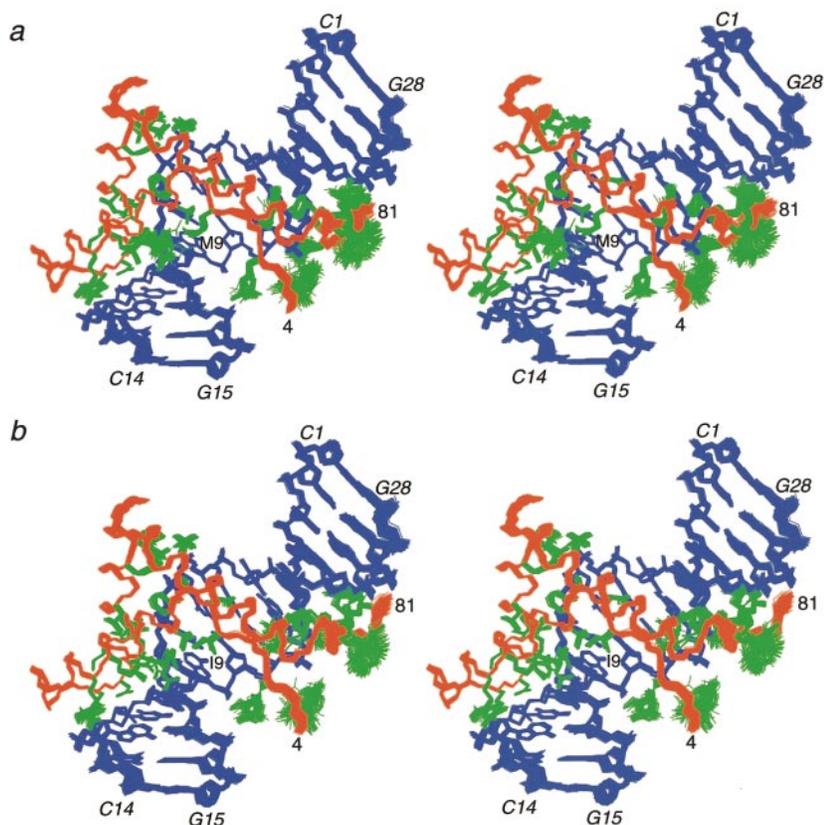


Figure 2. Stereoviews showing best-fit superpositions of the final 400 simulated annealing structures of (a) the wild-type and (b) the M9I mutant hSRY_{HMG}-DNA complexes. The protein backbone is shown in red, selected protein side-chains in green and the DNA in blue. The structures are best-fit to the backbone of residues 4-81 of the protein and base-pairs 1-14 of the DNA. Residues 1-3 at the N terminus and 82-85 at the C terminus are disordered. Labels for the DNA are in italics.

Overall architecture of the hSRY_{HMG}-complexes

The overall features of the wild-type and M9I complexes are also very similar (Figures 2 and 3(a)). The protein binds exclusively in the minor groove, resulting in large deformations in the DNA, which include bending, localized underwinding of the DNA double helix, expansion of the minor groove, and contraction of the major groove. Details of the DNA structure will be discussed in the next section. The DNA is nestled into the concave binding surface of the protein, and the protein contacts nine nucleotides in each strand,

with a two nucleotide offset between the strands (Figure 1(c)). The total accessible surface area buried upon complex formation is $\sim 3000 \text{ \AA}^2$ for both complexes. (The excluded accessible surface area can be decomposed into $\sim 1550/1580 \text{ \AA}^2$ and $\sim 1450/1420 \text{ \AA}^2$ for the protein and DNA moieties, respectively, of the wild-type/M9I complexes). A summary of the contacts is provided in Figure 1(c) and stereoviews illustrating the protein-DNA contacts are shown in Figure 5. There is a very large number of contacts with the DNA involving $\sim 40\%$ of the residues of hSRY_{HMG}: five in the N-terminal tail, six in helix 1, one in the loop connecting

Figure 1. Intermolecular contacts between hSRY_{HMG} and DNA. (a) and (b) Strips from 3D ^{13}C -separated/ ^{12}C -filtered NOE spectra illustrating NOE contacts between protons attached to ^{13}C of the protein and protons attached to ^{12}C of the DNA; intermolecular NOEs involving Met9 of wild-type hSRY_{HMG} and Ile9 of M9I hSRY_{HMG} are shown in (a) and (b) respectively. (c) A diagram illustrating the intermolecular contacts between wild-type and mutant hSRY_{HMG} and DNA. All the contacts are exclusively in the minor groove of the DNA. Contacts involving Met9 of wild-type hSRY_{HMG} and Ile9 of M9I hSRY_{HMG} are shown in red and blue, respectively; purple lines indicate potential salt-bridges or hydrogen bonding interactions. Excluding Met9/Ile9, the contacts observed in the wild-type and M9I complexes are essentially identical, with the possible exception of a potential hydrogen bond between the guanidino group of Arg78 and the N-3 atom of A26 in the M9I complex, which is indicated by a broken purple line. The sugars labeled in green (A8, T19 and T20) have a C-3'-endo sugar pucker, while the remainder are in the C-2'-endo/C-1'-exo conformation. The phosphate groups colored in orange have ^{31}P shifts in the range -3.0 to -3.5 ppm and are located at the top (C25) and bottom (A12) of the DNA-binding site; the phosphate groups colored in yellow (A9, T21 and G22) have ^{31}P shifts in the range -5.1 to -5.8 ppm and are located close to the center of the bend and the site of partial intercalation of Ile13 between base-pairs 8 and 9; the remaining phosphate groups (with the exception of the phosphate group of T20, which could not be observed) have ^{31}P shifts in the range -4.0 to -4.9 ppm. The ^{31}P shifts of A12, C25, G22, T21 and A9 are $-3.01/-3.08$ ppm, $-3.42/-3.51$ ppm, $-5.13/-5.13$ ppm, $-5.37/-5.34$ ppm and $-5.81/-5.79$ ppm, respectively, for the wild-type and M9I hSRY-DNA complexes.

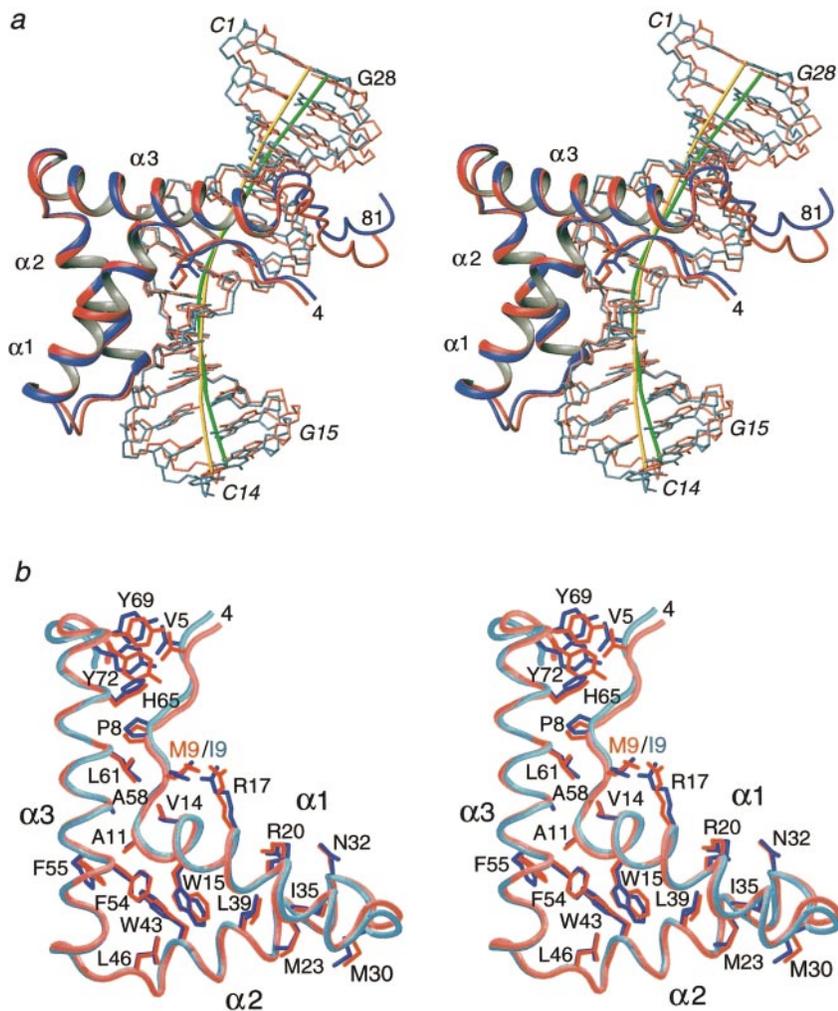


Figure 3. Stereoviews showing a best-fit superposition of the restrained regularized mean structures of the wild-type (red) and M9I (blue) hSRY_{HMG}-DNA complexes, illustrating (a) differential DNA bending and (b) packing interactions within the hydrophobic core. The protein backbone is shown as a ribbon, and the structures are best-fit to the backbone of residues 11-26 and 32-68 of the protein only. Also shown in (a) are the DNA helical axes for the wild-type and M9I complexes in green and yellow, respectively. Note that in wild-type hSRY_{HMG} there is a hydrogen bond between the guanidino group of Arg17 and the S^γ atom of Met9, which is absent from the M9I mutant, since Ile does not possess any appropriate hydrogen bonding acceptor group. Labels for the DNA are in italics.

helices 1 and 2, eight in helix 2, four in helix 3 and six in the C-terminal tail. Eleven residues contact the DNA bases in the wild-type (Arg7, Asn10, Phe12, Ile13, Arg20, Asn32, Ser33, Ser36, Tyr74, Pro76 and Arg78) with an additional residue (Ile9) also contacting DNA bases in the M9I complex. The DNA bend is centered around base-pairs 8 and 9, and is directed away from the body of the protein by a hydrophobic wedge consisting of Met9/Ile9, Phe12, Ile13 (which is partially intercalated between these two base-pairs) and Trp43 (Figure 5(b)).

Eight residues are involved in electrostatic interactions with functional groups of the bases. There are six potential direct hydrogen bonds, as judged by a distance of <3.5 Å between donor and acceptor heavy atoms in the ensemble of simulated annealing structures: between the hydroxyl group of Tyr74 and the N-3 atom of A6, and the guanidino group of Arg7 and the O-2 atom of T23 (Figure 5(a)); between the carboxamide group of Asn10 and the N-3 atom of G22 (Figure 5(b)); and between the carboxamide group of Asn32 and the N-3 atom of A10, the hydroxyl group of S33 and the N-3 atom of G18, and the hydroxyl group of

S36 and the O-2 atom of T19 (Figure 5(c)). There are also several potential water-bridged hydrogen bonding interactions (with a distance between donor and acceptor heavy atoms between 3.5 and 6 Å): between the guanidino group of Arg78 and the N-3 atom of G4 (and possibly the N-3 atom of A26 as well in the M9I complex) (Figure 5(a)); and between the guanidino group of Arg20 and the N-3 atoms of A9 and A10 (Figure 5(c)). In addition there is a large number of electrostatic interactions neutralizing eight phosphate groups per DNA strand. The majority of these involves Lys and Arg residues (Figure 1(c)).

DNA bending in the hSRY_{HMG} complexes

The overall structural features of the 14mer DNA in the wild-type and M9I complexes are clearly similar (Figures 2 and 3(a)). Excluding base-pair steps 8-10 (i.e. base-pairs 8/9, 9/10 and 10/11), the average helical twist and rise are 34.5(±2.3)° (Figure 6(c)) and 3.6(±0.2) Å, respectively. Base-pair steps 8-10, which include the site of partial intercalation of Ile13 at base-pair step 8, are severely underwound with an average helical

Table 1. Structural statistics

	wt hSRY _{HMC} -DNA		M9I hSRY _{HMC} -DNA	
	(\overline{SA}_{wt})	(\overline{SA}_{wt}) _r	(\overline{SA}_{M9I})	(\overline{SA}_{M9I}) _r
A. Experimental restraints				
rms deviations from experimental restraints				
Distance restraints (Å) (1755/1693) ^a	0.04 ± 0.00	0.04	0.03 ± 0.00	0.03
Torsion angle restraints (°) (433/429) ^a	0.29 ± 0.03	0.34	0.30 ± 0.03	0.25
³ J _{H_{Nz}} coupling restraints (Hz) (70/66)	0.81 ± 0.02	0.84	0.89 ± 0.02	0.90
¹³ C ^α / ¹³ C ^β shift restraints (ppm) (165/165)	1.01 ± 0.01	0.99	0.95 ± 0.01	0.95
DNA D_{HH} dipolar couplings (Hz) (55/53) ^b	0.58 ± 0.01	0.56	0.61 ± 0.05	0.75
Fixed distance heteronuclear dipolar coupling <i>R</i> -factors (%) ^c				
Protein ¹ D _{NH} (71/66)	5.5 ± 0.2	5.5	7.4 ± 0.1	7.6
Protein ¹ D _{CH} (67/67)	5.6 ± 0.3	6.3	9.9 ± 0.1	10.0
Protein ¹ D _{NC'} (68/62)	20.4 ± 1.3	18.9	29.1 ± 0.3	28.9
Protein ² D _{HNC'} (68/62)	19.4 ± 0.3	18.8	22.0 ± 0.4	21.6
DNA ¹ D _{NH} (9/10)	10.0 ± 0.2	10.2	16.1 ± 0.5	16.1
DNA ¹ D _{CH} (37/33)	13.4 ± 0.1	11.2	10.7 ± 0.3	10.7
B. Measures of structure quality				
rms deviations from idealized covalent geometry				
Bonds (Å)	0.003 ± 0	0.004	0.003 ± 0	0.004
Angles (deg.)	0.76 ± 0.01	0.81	0.76 ± 0.01	0.80
Improper torsions (deg.)	0.60 ± 0.03	0.79	0.59 ± 0.03	0.79
% Residues in most favorable region of Ramachandran plot ^d				
	95.6 ± 0.9	94.7	93.9 ± 1.2	94.7
Number of bad contacts per 100 residues ^d				
	5.1 ± 1.4	4.1	5.8 ± 1.5	5.9
C. Coordinate precision (Å)^e				
Protein backbone + DNA heavy atoms	0.23 ± 0.06		0.23 ± 0.07	
Protein heavy atoms + DNA heavy atoms	0.56 ± 0.04		0.56 ± 0.05	
Protein backbone	0.25 ± 0.08		0.25 ± 0.08	
Protein heavy atoms	0.73 ± 0.05		0.73 ± 0.05	
DNA heavy atoms	0.21 ± 0.07		0.22 ± 0.08	

The notation of the NMR structures is as follows: (\overline{SA}_{wt}) and (\overline{SA}_{M9I}) are the final 400 simulated annealing structures for the wild-type and M9I complexes, respectively; \overline{SA}_{wt} and \overline{SA}_{M9I} are the mean structures obtained by averaging the coordinates of the individual structures best-fit to each other (with respect to the backbone atoms of residues 4-81 of the protein, and base-pairs 1-14 of the DNA); (\overline{SA}_{wt})_r and (\overline{SA}_{M9I})_r are the corresponding restrained regularized mean structures. The number of terms for the various restraints is given in parentheses, the first number corresponding to the wild-type complex, the second to the M9I complex. None of the structures exhibited interproton distance violations >0.5 Å or torsion angle violations >5°.

^a The distance restraints comprise the following (wild-type/M9I mutant): 1130/1060 NOE-derived interproton distance restraints within the protein subdivided into 378/372 intraresidue and 340/304 sequential ($|i - j| = 1$), 280/253 medium-range ($1 < |i - j| \leq 5$) and 132/131 long-range ($|i - j| > 5$) interresidue restraints; 268/276 NOE-derived interproton distance restraints within the DNA, subdivided into 90/111 intraresidue, 161/150 sequential intrastrand and 17/15 interstrand restraints; 168/168 NOE-derived intermolecular interproton distance restraints; 72/72 distance restraints for 36 backbone hydrogen bonds located in helices and added during the final stages of refinement according to standard criteria;⁵⁸ 106/106 distance restraints for Watson-Crick hydrogen bonds within the DNA including distance restraints to prevent unduly large shearing of the base-pairs;⁷⁶ and 11/11 ambiguous distance restraints, represented by a $(\Sigma r^{-6})^{-1/6}$ sum with an upper bound of 6.5 Å, for potential salt-bridges between Lys and Arg side-chains of the protein and the phosphate groups of the DNA, which were added in the final stages of refinement according to the criteria proposed by Omichinski *et al.*⁶⁷ (The latter were as follows: from the guanidino group of Arg4, Arg17, Arg21 and Arg31 to the phosphorus atoms of C7/A8, A10, A11, and A13/A14, respectively, and from the N^δ atom of Lys6, Lys 37, Lys44, Lys51, Lys73, Lys79 and Lys81 to the phosphorus atoms of A8/A9, T19/T20, T21, G22, C25/A26, A6/C7 and C5/A6, respectively). There were 294/290 torsion angle restraints within the protein comprising 84/84 ϕ , 83/83 ψ , 68/68 χ_1 , 47/47 χ_2 and 12/8 χ_3 restraints; and 134/134 loose torsion angle restraints within the DNA (see Materials and Methods).

^b The homonuclear D_{HH} residual dipolar couplings were classified into approximate ranges corresponding to strong, medium and weak cross-peak intensities in a ¹³C-filtered correlated spectroscopy (COSY) spectrum and represented by square-well potentials in the target function. Consequently, only an rms deviation between observed upper and lower bounds and calculated D_{HH} couplings can be determined. Note that the sign of D_{HH} was not determined, so absolute values were employed.⁶⁵

^c The dipolar coupling *R*-factor is defined as the ratio of the rms deviation between observed and calculated values to the expected rms deviation if the vectors were oriented randomly. The latter is given by $\{2D_a^2[4 + 3\eta^2]/5\}^{1/2}$ where D_a is the magnitude of the alignment tensor and η is the rhombicity.³⁷ The orientation of the alignment tensor in the two complexes differs by less than 5°.

^d The percentage residues in the most favorable region of the Ramachandran plot and the number of bad contacts per 100 residues (which provides a measure of the quality of the non-bonded contacts) were determined using the program PROCHECK.⁸⁶ There are no ϕ , ψ angles in the disallowed region of the Ramachandran plot and the number of bad contacts per 100 residues lies inside the range observed for high-resolution (≤ 2 Å) crystal structures. The dihedral angle *G*-factors for ϕ/ψ , χ_1/χ_2 , χ_1 and χ_3/χ_4 are 0.34 ± 0.02 , 0.42 ± 0.07 , -0.018 ± 0.13 and -0.21 ± 0.15 , respectively for the wild-type complex, and 0.31 ± 0.02 , 0.35 ± 0.07 , 0.02 ± 0.15 and -0.27 ± 0.15 , respectively, for the M9I complex.

^e The precision of the coordinates is defined as the average atomic rms deviation between the individual 400 simulated annealing structures and the corresponding mean coordinates. The values refer to residues 4-81 of the protein and base-pairs 1-14 of the DNA. Residues 1-3 and 82-64 of the protein make no contacts with the DNA and are disordered in solution.

twist of 25.4(±2.4)° (Figure 6(c)), a helical rise of 3.3-3.4 Å for base-pair step 8 and 4.0-4.7 Å for base-pair steps 9 and 10. The minor groove for

base-pairs 7-11 is expanded and shallow to accommodate the protein, with a width of 11-12 Å and a depth of 0-2 Å. Concomitantly, the major groove

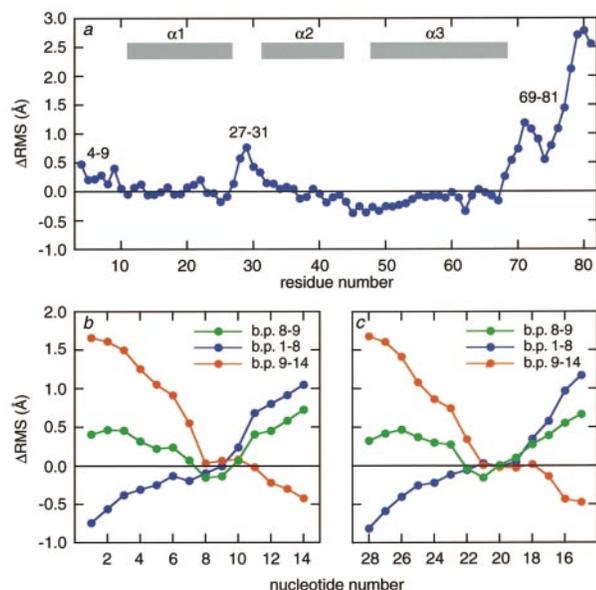


Figure 4. Quantitative description of the differences between the wild-type and M9I hSRY_{HMG}-DNA complexes plotted as Δ RMS versus residue number for (a) the protein backbone (N, C α , C and O atoms), (b) strand 1 of the DNA (bases 1-14) and (c) strand 2 of the DNA (bases 15-28). Δ RMS is given by the r.m.s. difference between the two restrained regularized mean coordinates minus the sum of the precision of the coordinates for the two ensembles of simulated annealing structures. Best-fitting in (a) is carried out with respect to the backbone of residues 11-26 and 32-68 of the protein; best-fitting in (b) and (c) is carried out with respect to base-pairs 1-8 (blue), base-pairs 9-14 (red) and base-pairs 8 and 9 (green). Positive values of Δ RMS indicate regions of real difference between the two sets of coordinates (i.e. regions where the two bundles of simulated annealing structures do not overlap).

for base-pairs 7-11 is compressed and deep with a width of 9-10 Å and a depth of 8-10 Å.

The sugar puckers for all nucleotides with the exception of A8, T19 and T20, are C-2'-endo/C-1'-exo with an average pseudo-rotation angle of 152(±15)°. A8, T19 and T20 are C-3'-endo with a pseudo-rotation angle of 36(±20)° (Figure 7(a)-(c)). The average sugar pucker amplitude is 28(±5)°.

The overall bend angle (averaged over 400 simulated annealing structures) for the 14mer in the wild-type complex is 54.0(±1.7)° versus 40.9(±1.9)° in the M9I complex (Figures 3(a) and 7). The difference in bend angle is evident when the complexes are best-fit to the protein only (backbone of residues 10-26 and 32-68) (Figure 3(a)), to base-pairs 1-8 (Figure 7(a)), to base-pairs 9-14 (Figure 7(b)), or to base-pairs 8 and 9 (Figure 7(c)). While small, the 13° difference is statistically significant, as is evident from the distribution of bend angles in the two ensembles of 400 simulated annealing structures, which display no overlap (Figure 7(d)).

The difference in DNA bending can be localized clearly between base-pairs 8 and 9, which is the site of both contact with the residue at position 9 (Met or Ile) and partial intercalation of Ile13. The overall atomic r.m.s. difference for base-pairs 1-14 in the two complexes is 1.0 Å, a value much larger than the errors in the DNA coordinates (Table 1). However, a best-fit to base-pairs 1-8 results in an atomic r.m.s. difference of only 0.33 Å for base-pairs 1-8, well within the errors of the coordinates (Figure 4(b) and (c), blue line), while base-pairs 9-14 are displaced by 2.3 Å. Similarly, a best-fit to base-pairs 9-14 results in an atomic r.m.s. difference of 0.55 Å for base-pairs 9-14, again within the errors of the coordinates (Figure 4(b) and (c), red line), with a corresponding displacement of 3.2 Å for base-pairs 1-8. Finally, even when the base-pairs 8 and 9 are best-fit (r.m.s. difference of 0.38 Å), the displacements of base-pairs 1-7 (1.5 Å) and 10-14 (1.7 Å) are statistically significant (Figure 4(b) and (c), green line).

The underlying origin of the difference in bend angles in the wild-type and M9I complexes resides in the nature of the residue at position 9: Met9 versus Ile9. A comparison of the interactions of Met9 and Ile9 with A8 and A9 is provided in Figures 8 and 9.

In the wild-type complex, Met9 contacts the sugar moieties of A8 and A9 (Figures 8 and 9(a)). Although a methionine side-chain possesses an intrinsic conformational flexibility on account of its linearity and length, the orientation and conformation of Met9 in the wild-type complex is stabilized by a hydrogen bond between the guanidino group of Arg17 and the S δ of Met9 (Figures 3(b) and 9(a)). As a result, the S δ atom of Met9 is in van der Waals contact with the sugar of A9 with distances to the O-4', C-4' and C-5' atoms of 4.2, 3.5 and 3.7 Å, respectively (Figures 8(e) and 9(a)).

In the M9I complex, Met9 is replaced by the "shorter" isoleucine residue, whose range of accessible side-chain torsion angles is severely limited by the presence of a β branch. The hydrogen bond involving Arg17 and Met9 in the wild-type complex is no longer possible in the M9I mutant. Given the very close similarity in the other intermolecular contacts and buried surface area, the absence of this hydrogen bond in the M9I mutant may contribute, in part, to the threefold weaker specific DNA binding (corresponding to a difference in the free energy of binding of only 0.6-0.7 kcal mol⁻¹ (1 cal = 4.18 J)) of M9I hSRY_{HMG} relative to wild-type. The C'm methyl group of Ile9 is packed tightly against the sugar moieties of both A8 and A9, directly contacting the H2'' proton of A8 and forming a putative methyl-oxygen hydrogen bond⁴¹ with the O-4' atom of A9 (Figures 8(e) and 9(b)). Thus, the distance from the C'(Met9)/C'm(Ile9) atom to the C-2' atom of A8 is reduced by ~1 Å (from 4.9 Å in the wild-type complex to ~4.0 Å in the M9I complex) (Figure 8(e)). The more intimate packing of the side-chain of Ile9 against the sugar moiety of A8

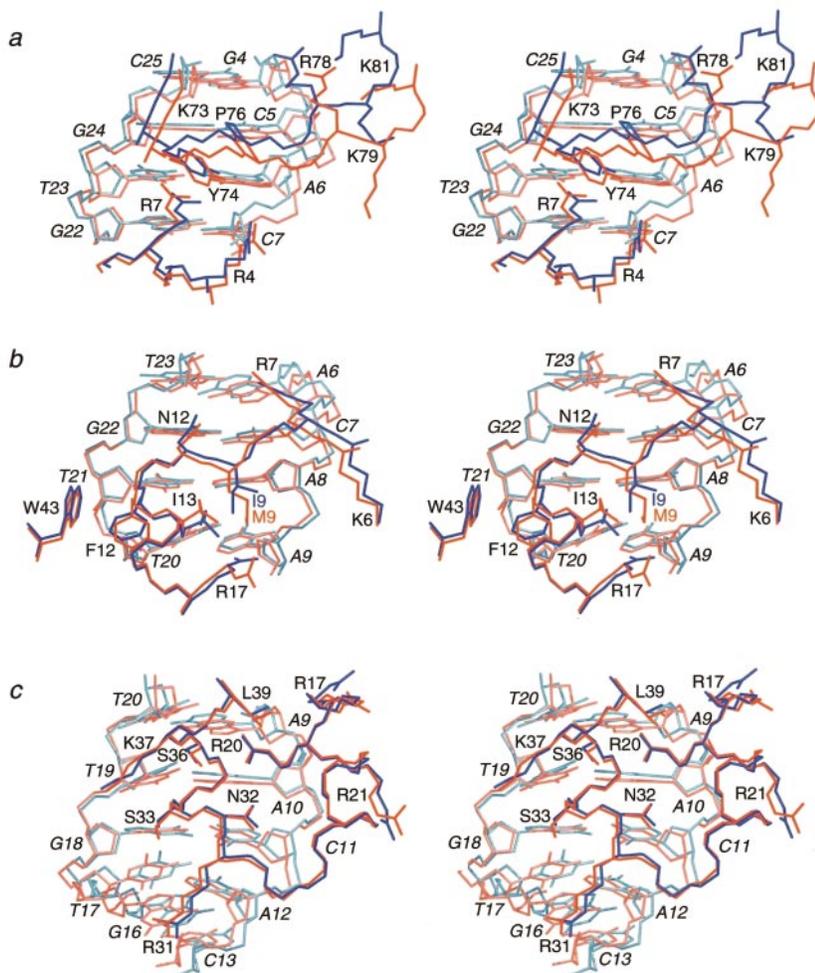


Figure 5. Stereoviews providing a comparison of the interactions between hSRY_{HMG} and DNA for the wild-type and M9I complexes involving (a) base-pairs 4-7, (b) base-pairs 6-9 and (c) base-pairs 9-13. All superpositions involve the restrained regularized mean structures and best-fitting is carried out using the backbone of residues 11-26 and 32-68 of the protein only. The wild-type and M9I complexes are depicted in red and blue, respectively, with varying shades for the side-chains (darkest), protein backbone (intermediate) and DNA (lightest). The view shown in (b) includes the site of partial intercalation of Ile13 between base-pairs 8 and 9, and the site of the mutation (Met9 in wild-type and Ile9 in the mutant). Labels for the DNA are in italics.

relative to that of Met9 can also be readily appreciated from the surface representations shown in Figure 8(a)(d) and the space-filling models shown in Figure 9(a) and (b). The distances from the C^γ of Met9 and C^γm of Ile9 to the O-4', C-4' and C-5' atoms of A9 are comparable (4.5 Å *versus* 4.3 Å, 4.2 Å *versus* 4.3 Å, and 3.8 Å *versus* 3.9 Å, respectively).

Thus, in effect, the bulky S^δ of Met9 in the wild-type complex pushes the sugar moiety of A9 away, thereby increasing the roll and tilt between base-pairs 8 and 9 relative to that in the M9I complex (Figures 6(a) and (b) and 8(e)). This, in turn, permits slightly deeper intercalation of Ile13 in the wild-type complex (Figure 8(a)) relative to the M9I complex (Figure 8(c)). Ile9, on the other hand, functions as a brace between the sugar moieties of A8 and A9. This description is highlighted by chimeras constructed from a best-fit to the base of A8 comprising wild-type hSRY_{HMG} and DNA from the M9I complex (Figure 9(c)), and M9I hSRY_{HMG} and DNA from the wild-type complex (Figure 9(d)). In the wild-type hSRY_{HMG}-mutant DNA chimera, there is a steric clash (i.e. atomic overlap) between the S^δ (and to a lesser extent the C^γ) of Met9 and the sugar moiety of A9 (Figure 9(c)). In particular,

the partially negatively charged S^δ atom of Met9 and O-4' atom of A9 are in repulsive contact. In the converse M9I hSRY_{HMG}-wild-type DNA chimera, the packing of Ile9 against the sugar moiety of A9 is clearly suboptimal (Figure 9(d)). In short, substitution of Met9 by Ile results in significant changes in DNA bending that are clearly detectable in the NMR structures.

Base-pair roll and tilt as the origin of DNA bending in the hSRY_{HMG}-DNA complexes

A summary of the variations in global base-pair roll, tilt and helical twist as a function of base-pair step is shown in Figure 6. Crystal structures of B-DNA are generally anisotropic with respect to roll and tilt:^{42,43} that is, the variations in roll are much greater than those in tilt. While this is true in the M9I complex, where the variation in tilt (~10°) is half that in roll (~20°), this is less so for the wild-type complex, where both roll and tilt vary over a range of ~20° (Figure 6(a) and (b)). The two base-pair steps with the largest tilt angles (Figure 6(a)) have one or two sugar moieties in the C-3'-*endo* conformation (Figure 1(c)): the base-pair steps are 8 and 10, and the C-3'-*endo* sugar moieties are

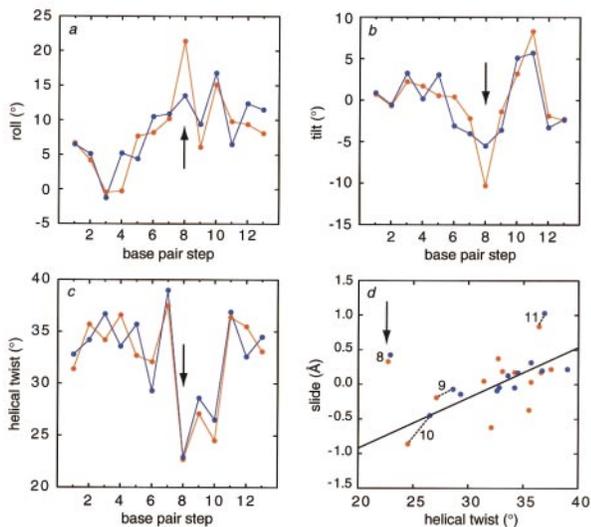


Figure 6. Comparison of base-pair (a) roll, (b) tilt and (c) twist for the DNA in the wild-type (red) and M9I (blue) hSRY_{HMG}-DNA complexes. Also shown in (d) is a plot of base-pair slide *versus* helical twist. The arrow indicates the site of partial intercalation of Ile13 at base-pair step 8 (i.e. between base-pairs 8 and 9). Note that (c) base-pair steps 8, 9 and 10 are highly underwound and (d) that base-pair step 8 deviates from the linear slide-twist correlation.

located in base-pairs 8 (A8), 9 (T20) and 10 (T19). The switch in sugar pucker conformation from C-2'-*endo* to C-3'-*endo* decreases the sequential P-P distance by ~ 1 Å,⁴⁴ thus inducing asymmetry in the DNA backbone. This results in a “mixed” tilt-roll local bend, similar to that for the B/A junction, first predicted in computations^{43,45} and subsequently observed crystallographically.⁴⁶ The structures of the DNA in both the wild-type and M9I complexes are entirely consistent with these observations: the C-2'-*endo* sugar moieties produce P(*i*)-P(*i* + 1) distances of 6.5-7.1 Å (average 6.8(±0.1) Å), while the C-3'-*endo* sugar moieties (at base position *i*) have P(*i*)-P(*i* + 1) distances of 5.9-6.2 Å (average 6.1(±0.1) Å). The decreased P(*i*)-P(*i* + 1) distances occur at base-pair step 8 (both strands) and base-pair step 9 (second strand).

Overall, the pattern of C-3'-*endo* sugar moieties seen in the 14mer DNA bound to hSRY_{HMG} is consistent with the pattern predicted for A/B junctions by Selsing *et al.*:⁴⁵ namely, the positions of the C-2'-*endo* and C-3'-*endo* sugar moieties comply with the following rule: going along a DNA strand in the 5' to 3' direction, the switch in sugar pucker conformation (either from C-3'-*endo* to C-2'-*endo* or *vice versa*) occurs first in the selected strand, and then in the opposite strand. Thus, we have the pattern:

5'... -B-B-A-B-B-B-...

3'... -B-B-B-A-A-B-B-...

where B and A denote C-2'- and C-3'-*endo* sugar puckers, respectively. The two A/B junctions would be expected to produce a mixed tilt-roll

bend directed towards the major groove and one of the strands, resulting in a zig-zagged structure with a net bend towards the major groove. This is exactly what is observed in the 14mer DNA bound to hSRY_{HMG}: the center of the bend is localized close to base-pair 9 (at the center of the C-3'-*endo* patch) and is directed into the major groove.

The general picture presented above is consistent with the roll profile (Figure 6(a)). The highest roll angles are observed in base-pair steps 8 and 10; that is, the steps containing nucleotides with the C-3'-*endo* sugar moieties. These roll angles taken together result in a cumulative major groove bend centered at base-pair 9. Thus, both descriptions, the general one in terms of A/B junctions, and the detailed one in terms of roll angles, yield the same picture.

The helical twist angles between the wild-type and M9I complexes differ by 3-4°, mostly at the ends of the duplex (Figure 6(c)), and this modest difference does not produce any significant change in the rotational orientation of the local bends. Thus, the overall effect of the difference in twist angles between the two complexes is negligible. Note that for both complexes, the twist angles are unusually low for base-pair steps 8-10 (Figure 6(c)), and correlate nicely with the C-3'-*endo* sugar pucker seen in these base-pairs. The twist-slide correlation follows the trend observed in DNA crystal structures⁴³ for all dimeric base-pair steps with the exception of base-pair steps 8 and 11 (Figure 6(d)). This is apparently a reflection of the large tilt angles at these two base-pair steps (cf. Figure 6(b)).

The maximal difference between the DNA in the wild-type and M9I complexes is located at base-pair step 8: $\sim 8^\circ$ in roll and $\sim 5^\circ$ in tilt. The effects of roll and tilt are cumulative, since they are located at the same base-pair step, resulting in an overall difference in bend angle of $\sim 13^\circ$ between the two complexes. Although there are other small differences in roll and tilt angles at other base-pair steps, these cancel each other, as evidenced by the fact that the atomic r.m.s. differences between base-pairs 1-8 and 9-14 taken individually are small (0.33 and 0.55 Å, respectively) and within the coordinate errors (Figure 4(b) and (c)).

Comparison of the hSRY_{HMG}-DNA complex with other HMG box-DNA complexes

The construct of hSRY_{HMG} used in the present study makes it possible to delve further into the differences between the hSRY_{HMG}-DNA complex, and the complexes of other HMG-box domains, specific (LEF-1) and non-specific (HMGD and NHP6A), whose structures are known. To facilitate comparisons, the numbering schemes employed for both the protein and DNA components of these complexes will be those of the present hSRY_{HMG}-DNA complex.

The original structure of hSRY_{HMG} bound to DNA was unable to answer a number of questions regarding the interactions of the extended C-term-

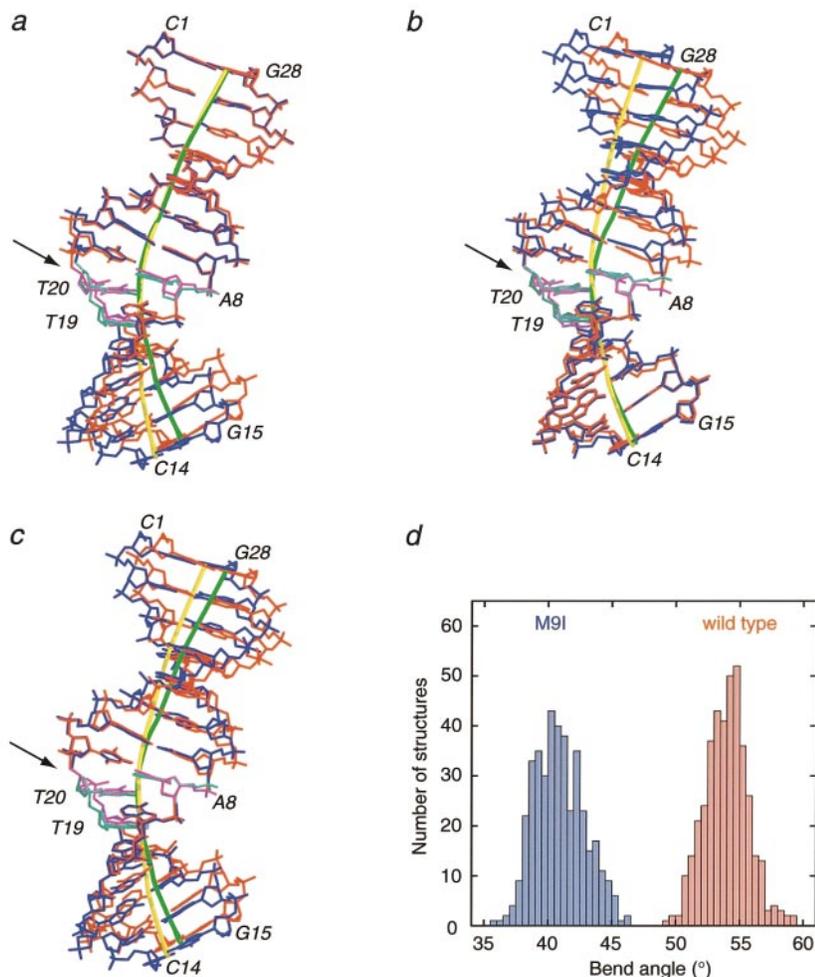


Figure 7. Differential DNA bending induced by wild-type and M9I hSRY_{HMG}. Superposition of the DNA of the restrained regularized mean structures of the wild-type (red) and M9I mutant (blue) complexes best-fit to (a) base-pairs 1-8, (b) base-pairs 9-14 and (c) base-pairs 8-9. The three nucleotides (A8, T19 and T20) with C-3'-endo sugar puckers are shown in magenta for the wild-type and cyan for the M9I mutant. The DNA helical axes for the wild-type and M9I complexes are shown in green and yellow, respectively. (d) Histogram illustrating the distribution of DNA bend angles (in 0.5° interval bins) observed in the ensemble of 400 simulated annealing structures for the wild-type (red) and M9I (blue) complexes. The average DNA bend angle is 54.0(±1.7)° for the wild-type complex and 40.9(±1.9)° for the M9I complex.

inal tail owing to the shorter length of the hSRY_{HMG} construct employed (residues 1-78). In the case of LEF-1,¹³ for example, interactions with the DNA extend out to residue 84 and the C-terminal tail wraps around and binds across the narrowed major groove, enhancing DNA bending (Figure 10(a) and (b)). The importance of these interactions for LEF-1 is further supported by deletion analysis and mutagenesis, which indicate that the C-terminal tail is an important contributor to both DNA-binding affinity and DNA bending.^{13,47,48} In a similar vein, the N-terminal tail of NHP6A wraps around the major groove in a manner mimicking the C-terminal tail of LEF-1.¹⁵ The present hSRY_{HMG} construct extends out to residue 85, and no interaction involving the major groove is observed either at the N terminus or the C terminus. hSRY_{HMG} and LEF-1¹³ are similar in structure, with a C^α backbone r.m.s difference of 1.4 Å for 67 residues (specifically, residues 4-27, 31-43 and 46-75; the sequence identity for these 67 residues is 28%), and the direction of the polypeptide chain deviates only beyond residue 75 (Figure 10(a)). A key player is Tyr74 in hSRY_{HMG} which hydrogen bonds to N-6 of A9. Tyr74 is substituted by Ala in LEF-1. The position of Tyr74 in hSRY_{HMG}, how-

ever, is approximately occupied by Tyr78 in LEF-1 (Figure 10(b)), an offset of four residues, which accounts, in part, for the change in direction of the polypeptide chain such that the C-terminal tail is directed along the minor groove in the hSRY_{HMG} complex but towards the major groove in the LEF-1 complex. The segment of polypeptide chain from residues 81-84 in LEF-1 is highly basic: Lys81, Lys82, Lys83 and Arg84 interact with the sugar-phosphate backbone of the major groove and their backbone positions are located approximately opposite the site of partial intercalation of Met13 (Figure 10(b)). These major groove interactions support the deeper intercalation of Met13 between base-pairs 8 and 9 relative to Ile13 in the hSRY_{HMG} complex, which in turn results in a substantial increase in roll at this base-pair step in the LEF-1 complex (~40°) relative to that in the wild-type hSRY_{HMG} complex (~20°, Figure 6(a)). In contrast, in hSRY_{HMG}, the residues at positions 82, 83 and 84 are substituted by neutral residues, namely Met82, Leu83 and Pro84, thereby effectively eliminating non-specific electrostatic interactions between this segment of the C-terminal tail and the DNA in the hSRY_{HMG} complex. These factors readily account for the higher degree of DNA bending observed in

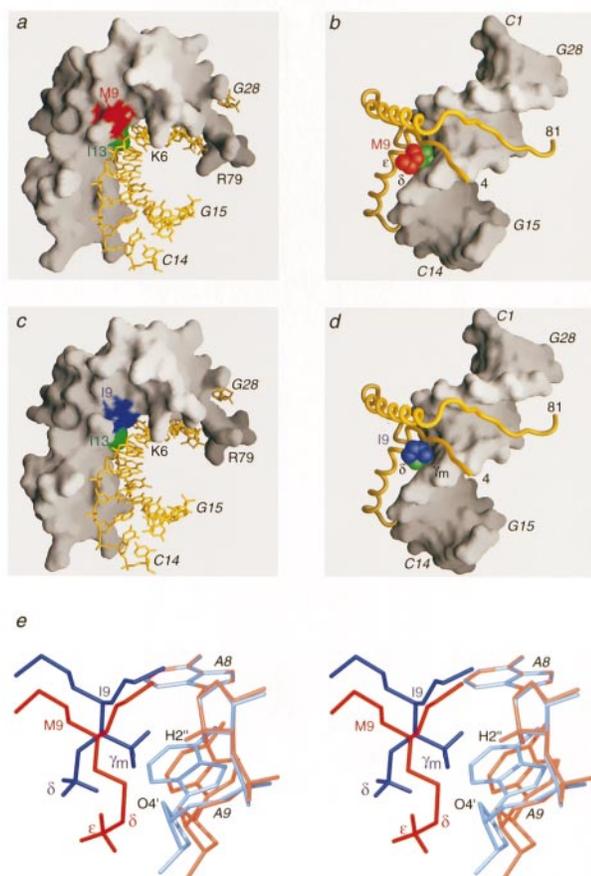


Figure 8. Comparison of the intermolecular contacts with DNA involving ((a), (b) and (e)) Met9 of wild-type hSRY_{HMG} and ((c), (d) and (e)) Ile9 of the M9I mutant. In (a) and (c) the DNA is displayed as a stick representation in yellow and the protein is shown as a molecular surface with the location of (a) Met9 indicated in red, (c) Ile9 in blue, and Ile13 in green (which is seen to partially intercalate between base-pairs 8 and 9). In (b) and (d) the DNA is displayed as a molecular surface, the protein backbone is shown as a yellow tube, and the side-chain atoms (as CPK models) of (b) Met9 are shown in red, (d) Ile9 in blue, and Ile13 in green. In (e) a stereoview of a superposition of the wild-type and M9I structures, best-fit to the base of A8 and illustrating a detailed view of the interactions of Met9 and Ile9 with the sugar moieties of A8 and A9, is shown with the wild-type in red and the mutant in blue (darker shades are used for the protein, lighter ones for the DNA). Protons are shown only for the methyl groups of Met9 and Ile9 and the H2'/H2'' protons of A8. The S^δ atom of Met9 in the wild-type complex is in van der Waals contact with the sugar moiety of A9: the distances from the S^δ atom to the O-4', C-4' and C-5' atoms of A9 are 4.2, 3.5 and 3.7 Å, respectively. The distances from the C^γ atom of Met9 in the wild-type to the C-2' atom of A8, and the O-4', C-4' and C-5' atoms of A9 are 4.9, 4.5, 4.2 and 3.8 Å, respectively; the corresponding distances involving the C_m^γ methyl group of Ile9 in the M9I mutant are 4.0, 4.3, 4.3 and 3.9 Å, respectively. Labels for the DNA are in italics.

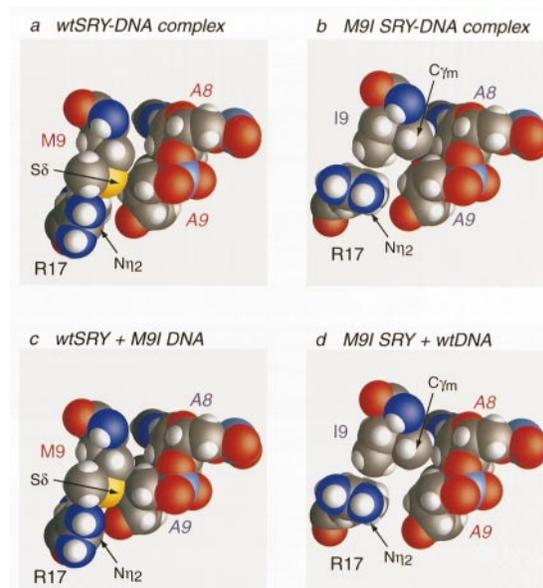


Figure 9. Space-filling models illustrating the interaction of Met9/Ile9 with Arg17, A8 and A9. (a) Wild-type hSRY_{HMG}-DNA complex; (b) M9I hSRY_{HMG}-DNA complex; (c) chimera of wild-type hSRY_{HMG} and DNA from M9I complex; (d) chimera of M9I hSRY_{HMG} and DNA from wild-type complex. The color coding of the atoms is as follows: H, light grey; C, dark grey; N, dark blue; O, red; S, yellow; and P, light blue. All structures are shown in the same view, best-fit to the base of A8.

the NMR structure of the LEF-1 complex ($\sim 75^\circ$; Love *et al.*¹³) relative to that of the wild-type hSRY_{HMG} complex ($\sim 55^\circ$).

In the crystal structure of the HMGD-DNA complex,¹⁴ the DNA is even more bent ($\sim 100^\circ$) (Figure 10(c)). The C^α backbone atomic r.m.s. difference between hSRY_{HMG} and HMGD is 1.1 Å for 64 residues (residues 5-47 and 50-70 with a two residue deletion in the loop connecting helices 2 and 3 in HMGD relative to hSRY_{HMG}; the sequence identity for these 64 residues is 22%). As in the case of LEF-1, the intercalating residue is Met13, which in the HMGD complex penetrates more deeply between base-pairs 8 and 9 than Ile13 of hSRY_{HMG}, resulting in a roll angle ($\sim 45^\circ$) comparable to that observed in the LEF-1 complex. In contrast to the LEF-1 complex, there is no C-terminal tail. Additional bending in the HMGD complex, however, occurs between base-pairs 10 and 11 with partial intercalation of Val32 resulting in an approximate doubling of the roll angle ($\sim 30^\circ$) at this base-pair step (Figure 10(d)) relative to the hSRY_{HMG}-DNA complex ($\sim 15^\circ$ for both the wild-type and M9I complexes; Figure 6(a)). Rather than intercalate, the equivalent residue in the hSRY_{HMG} complex, Asn32, is hydrogen bonded to the base of A10 (Figure 5(c)).

The interactions of hSRY_{HMG}, LEF-1 and HMGD with their target sites provide some general con-

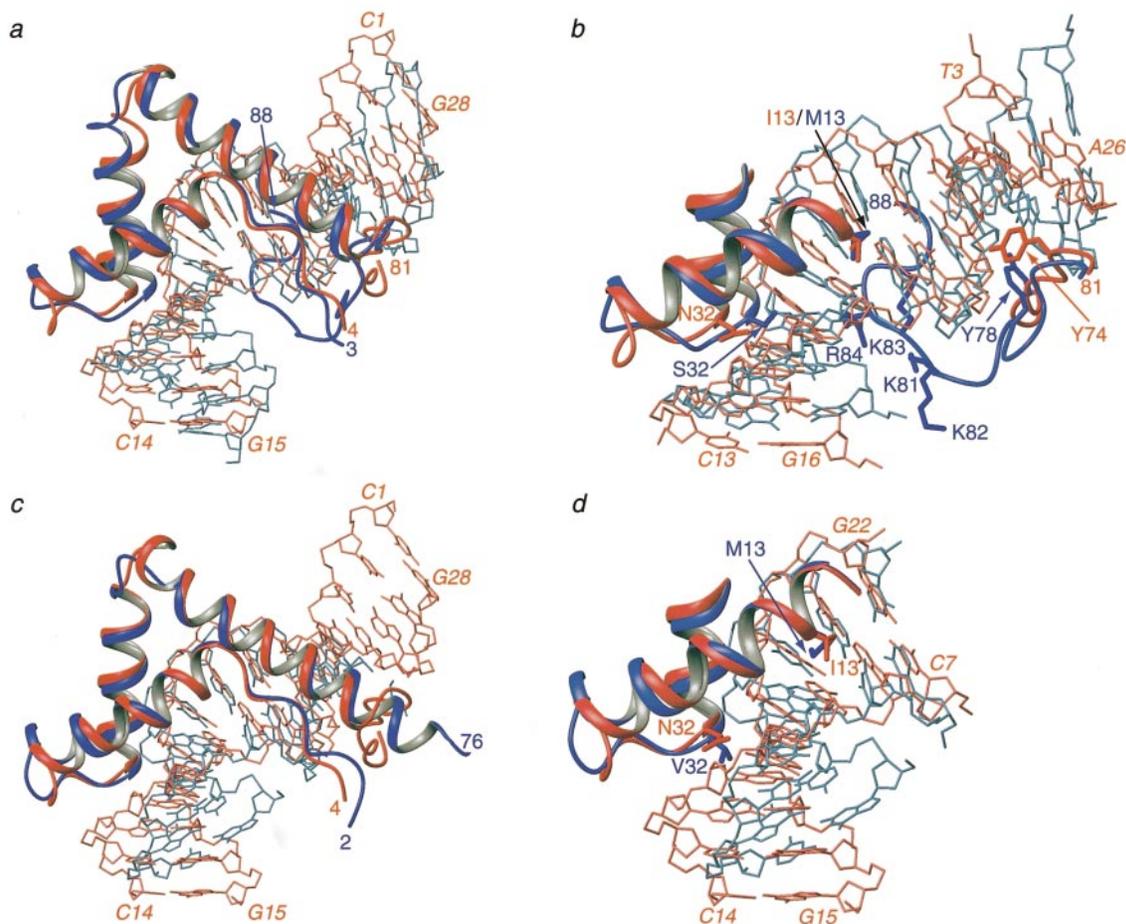


Figure 10. Comparison of the wild-type hSRY_{HMG}-DNA structure with the NMR structure of the LEF-1-DNA complex ((a) and (b)) and the crystal structure of the HMGD-DNA complex ((c) and (d)). The hSRY_{HMG} complex is shown in red, and the other two complexes in blue. The residue numbering employed for LEF-1 has been changed to correspond to that of the present hSRY_{HMG} construct; to obtain the original numbering of LEF-1, subtract 3 from the indicated residue numbers. Overall views are shown in (a) and (c), and more detailed views including relevant side-chains are shown in (b) and (d). Labels for the DNA are in italics. The coordinates of the LEF-1-DNA and HMGD-DNA complexes are taken from Love *et al.*¹³ (PDB code 2LEF) and Murphy *et al.*¹⁴ (PDB code 1QRV).

clusions with regard to the mechanism whereby these proteins bend DNA. The nature of the intercalating residue at position 13 is clearly important: Met13 in LEF-1 and HMGD is capable of deeper intercalation than Ile13 in hSRY_{HMG} and hence can result in a higher degree of bending. However, this is not the only factor, since the HMG-box of SOX9, which has 46% sequence identity with hSRY_{HMG}, has a methionine residue at position 13 and yet bends the DNA to the same extent as hSRY_{HMG} in circular permutation gel-shift assays.⁴⁹ Other factors clearly include the following: (i) the identity of residues 9, 12, 13 and 43, which make up the hydrophobic wedge around the site of intercalation; (ii) the identity of residue 32 at the bottom of the binding site and, in particular, whether this residue participates in either hydrogen bonding (as in the hSRY_{HMG} complex) or partial intercalation (as in the HMGD complex); and (iii) the nature of the C-terminal tail and whether it comprises a string of positively charged residues that facilitates

wrapping around the major groove (as in the LEF-1 complex). As discussed above, all these factors come into play in the case of the hSRY_{HMG}, LEF-1 and HMGD complexes.

Concluding Remarks

Here, we have provided a structural basis for understanding the effects of the M9I mutation (equivalent to M64I in full-length hSRY) upon DNA bending. Consistent with previous circular permutation gel-shift assays, the M9I mutation reduces DNA bending. Wild-type hSRY_{HMG} bends the DNA by $\sim 54^\circ$, on average, and the M9I mutant by $\sim 41^\circ$, a reduction of 13° in the average bend angle. The absolute values of the average bend angles derived from the present structures are somewhat smaller (by $\sim 15\text{--}20^\circ$) than those estimated by the circular permutation gel-shift assays ($\sim 75^\circ$ and $\sim 55^\circ$ for wild-type and M9I hSRY_{HMG}, respectively;³⁰). These quantitative differences

between the structural and biochemical results are not surprising, given the indirect nature of the gel-shift assays and the numerous underlying assumptions required to derive bend angles from the circular permutation gel-shift data.^{50,51} Our current structural results are consistent with results from electron microscopy on other systems,⁵² which have shown that the circular permutation gel-shift assay tends to systematically overestimate the bend angle.

The structures presented here demonstrate clearly how subtle changes in the interaction of a critical residue with the DNA, arising from a conservative mutation, can readily alter local base-pair spatial organization, perturbing roll and tilt angles, and thereby modulating the extent of DNA bending. Although a difference in bend angle of only 13° between the wild-type and M9I complexes may seem rather modest, it clearly has a rather dramatic effect *in vivo*, given that the M9I mutation in hSRY_{HMG} results in complete gonadal dysgenesis and 46X,Y sex reversal with full penetrance. However, this small difference in bend angle translates to large displacements over the sorts of distances involved in the formation of a transcription initiation complex. Thus, a 13° difference in bend angle would result in a ~75 Å displacement of the end of a 100 base-pair segment, assuming one arm of the loop induced by hSRY_{HMG} is fixed and the other arm is bent by either 54° in the wild-type complex or 41° in the M9I complex. (In the context of a static model, the effect of a difference $\Delta\theta$ in bend angle is equal to $2L\sin(\Delta\theta/2)$, where $L \sim 340$ Å). A DNA end displacement of this order of magnitude would clearly be more than sufficient to prevent distally placed proteins in the transcription initiation complex from interacting.

Finally, we note that the structural picture afforded by the NMR structures of the wild-type and M9I hSRY_{HMG} complexes is a static one. In a dynamic situation,⁵³ involving longer DNA fragments (100 base-pairs or more) where the overall bend angles may vary over a much wider range as a consequence of thermal fluctuations, the effects of the 13° difference in the average DNA bending observed in the NMR structures of the wild-type and M9I hSRY_{HMG} complexes would be further magnified.

Materials and Methods

Mutagenesis and Purification of hSRY_{HMG}

Wild-type hSRY_{HMG} and the M9I mutant, spanning residues 57-140 of full-length hSRY (and numbered residues 2-85 in the present construct), plus an N-terminal Met, were expressed using the PET21a vector and *Escherichia coli* BL21(DE3) cells. Thus, Met9 in the present construct is equivalent to Met64 in the full-length hSRY sequence. The original 78 residue clone (comprising residues 57-133 plus an N-terminal Met;¹²) was extended to 85 amino acid residues using an extended PCR primer. The M9I mutant was derived from the 85 residue wild-type hSRY_{HMG} construct using site-specific PCR muta-

genesis. The nucleotide sequence of both clones was confirmed by sequencing.

Cells were grown at 37°C in a modified minimal medium for uniform (>95%) ¹⁵N and/or ¹³C-labeling with ¹⁵NH₄Cl and [¹³C₆]glucose as the sole nitrogen and carbon sources, respectively. Cells were resuspended in lysis buffer (50 mM Tris, 10 mM EDTA, 10 mM benzamidine HCl, 10 mM DTT, pH 7.5) and lysed by passage through a French press, first at 500 psi (1 psi ≈ 6.9 kPa) and then a second pass at 700 psi, followed by sonication at 4°C. The soluble fraction was brought to 1 M NaCl and 5 ml of polyimine P/100 ml was added to pellet DNA. The supernatant was brought to 70% (w/v) ammonium sulfate to separate the proteins, and the resulting precipitate was suspended in lysis buffer, and dialyzed *versus* lysis buffer to remove salt. The resulting solution was passed over a P-cellulose column with a linear gradient to 90% of 1.5 M KCl in 10 mM sodium phosphate over 1440 ml, and the concentration of protein in the eluent was followed by UV spectroscopy. The peak was collected and dialyzed *versus* 10 mM sodium phosphate (pH 6.8). The dialysate was run over a preparative Mono-S column with a linear gradient from 30% to 75% of 1 M NaCl in 10 mM sodium phosphate over 500 ml and followed by UV spectroscopy. Fractions were collected, pooled, dialyzed, and characterized by mass spectroscopy.

Unlabeled 14-mer DNA, containing the specific SRY binding site within the *Amh* (also known as *Mis*) gene, was purchased from Midland Certified Reagent Company in single strands (5'-d-CCTGCACAAACACC and 5'-d-GGTGTTTGTGCAGG) that were HPLC-purified. The strands were annealed and then further purified on a preparative Mono-Q ion-exchange column with a linear gradient from 30% to 75% of 1 M NaCl in 10 mM sodium phosphate (pH 6.8), 1 mM EDTA over 180 ml and followed by UV spectroscopy. Fractions were collected, pooled, and dialyzed into 10 mM sodium phosphate (pH 6.8), 1 mM EDTA, 5 mM sodium azide. Double-stranded ¹³C/¹⁵N-labeled DNA was prepared and purified using the endonuclease sensitive repeat amplification (ESRA) procedure as described.⁵⁴ For technical reasons related to the ESRA method and the need for digestion with a restriction enzyme, the ¹³C/¹⁵N-labeled DNA employed was a 17mer,⁵⁴ where base-pairs 2-14 of the 14mer corresponded to base-pairs 4-16 of the 17mer. No differences in chemical shifts for wild-type or M9I hSRY_{HMG} or for the DNA corresponding to base-pairs 3-13 of the 14mer DNA was observed between the complexes with the 14mer and 17mer DNA. Since the majority of the work was carried out on the complexes with the 14mer, the structure calculations were carried out only for the 14mer complexes.

The 1:1 hSRY_{HMG}-DNA complexes were made by adding the protein (at low concentration) slowly with mixing to DNA (at low concentration). The complex was then concentrated and dialyzed against 10 mM sodium phosphate (pH 7.0), 1 mM EDTA, 5 mM sodium azide to yield final NMR samples containing approximately 1 mM 1:1 hSRY_{HMG}-DNA complex at pH 7.0.

Circular permutation electrophoretic mobility shift assays

The pBend 2 plasmid was obtained from Sankar Adhya.⁵⁵ The *Amh* binding site was inserted between the *Xba*I and *Sal*I sites in the plasmid, and the circular permutation region of the plasmid (between the *Eco*RI and

HindIII sites) was amplified using PCR. The PCR fragment was digested using *MluI*, *ClaI*, *EcoRV*, *KpnI* and *BamHI* to create a circularly permuted series. The protein and DNA fragments were added to a buffer solution containing 10 mM sodium phosphate, 1 mM EDTA, 100 mM NaCl, and 8% (w/v) Ficoll (as a stabilizing agent). The starting solutions and the resulting mixture were kept on ice before loading onto a 10% (w/v) polyacrylamide gel. The bands were stained with ethidium bromide and exposed with UV light for photography. Bend angles were calculated using the method of Thompson & Landy.⁵⁰

NMR spectroscopy

Spectra were recorded at 35 °C on Bruker DMX600, DMX500, DMX750 and DRX800 spectrometers. Each experiment was run on both complexes, wild-type and M9I mutant, using the same parameters for each sample with respect to both the data collection and analysis. All experiments were processed using the NMRPipe package,⁵⁶ and analyzed using PIPP, CAPP and STAPP.⁵⁷

Sequential assignment of protein ¹H, ¹⁵N, and ¹³C chemical shift assignments was achieved by means of through-bond heteronuclear scalar correlations along the protein backbone and side-chains^{31–33} using 3D CBCANH, CBCA(CO)NH, C(CCO)NH, H(CCO)NH, HBHA(CO)NH, HNHA, HCCH-COSY, HCCH-TOCSY, and CCH-COSY experiments. Assignment of ¹H, ¹⁵N and ¹³C DNA chemical shifts was obtained from analysis of 2D ¹²C-filtered NOE, ¹²C/¹⁴N-filtered NOE and ¹²C-filtered HOHAHA experiments recorded on a 1:1 complex of ¹⁵N/¹³C-hSRY_{HMG} (wild-type and M9I mutant) and unlabeled DNA, and from 2D ¹H-¹⁵N and ¹H-¹³C HSQC experiments and 3D ¹⁵N and ¹³C-separated NOE spectra recorded on a 1:1 complex of ¹⁵N-hSRY_{HMG} and ¹⁵N/¹³C-DNA, using conventional sequential assignment methodology for nucleic acids.⁵⁸ ³¹P assignments were obtained by correlation to the H3'(i – 1) and H4'(i) resonances observed in a 2D ¹H-³¹P HSQC spectrum.

Interproton distance restraints within the protein were obtained from 3D ¹³C and ¹⁵N-separated NOE spectra and from 4D ¹³C/¹³C- and ¹³C/¹⁵N-separated NOE spectra. Interproton distance restraints within the DNA were obtained from 2D ¹²C-filtered NOE spectra (using a range of mixing times to account for spin diffusion) recorded on a complex of ¹⁵N/¹³C-hSRY_{HMG} and unlabeled DNA, from 3D ¹⁵N and ¹³C-separated NOE spectra recorded on complexes of ¹⁵N-hSRY_{HMG} and ¹⁵N/¹³C-DNA. Intermolecular interproton distance restraints were obtained primarily from 3D ¹³C-separated/¹²C-filtered NOE experiments.

³J_{H_Nz}, ³J_{NC'} (aromatic, methyl and methylene), ³J_{CC'} (aromatic, methyl and methylene), ³J_{CC'} and ³J_{CαCδ} scalar couplings were measured by quantitative *J* correlation spectroscopy.⁵⁹ Backbone φ and φ torsion angle restraints were derived from ³J_{H_Nz} and ³J_{CC'} coupling constants in combination with a database search procedure based on backbone (N, C', C^α, C^β, H^α) chemical shifts using the program TALOS.⁶⁰ Side-chain torsion angle restraints were derived from analysis of the NOE/ROE and three-bond heteronuclear scalar coupling data.

³J_{H3'-P} couplings for the DNA, which are related to the ε (C4'-C3'-O3'-P) sugar-phosphate backbone angles, were also obtained by quantitative *J* correlation spectroscopy using a ¹²C-filtered constant time ¹H-¹H{³¹P} COSY difference experiment.⁶¹ Restraints for the δ (C5'-C4'-C3'-

O3') torsion angle, which is related to the sugar pucker, were deduced from a combination of a very short mixing time (20 ms) ¹²C-filtered NOE spectrum, a ¹²C-filtered rotating frame Overhauser effect (ROE) spectrum (mixing time 40 ms), and an out (H1' and H3') and back (H2'/H2'') ¹²C-filtered COSY spectrum (G.M.C. & A. Bax, unpublished results). The latter provides information on the relative magnitudes of the ³J_{H1'H2'} to ³J_{H1'-H2''} and the ³J_{H3'H2'} to ³J_{H3'-H2''} couplings. All the sugar moieties were found to be in the C-2'-endo range with the exception of those of A8, C14, T19 and T20, which were in the C-3'-endo range.

Residual heteronuclear dipolar couplings corresponding to fixed length vectors were determined from the difference in corresponding *J* splittings measured in a liquid crystalline medium of bicelles (4.5 to 5% 3:1 DMPC:DHPC;³⁴) and in isotropic (water) medium. Dipolar couplings for the protein (¹D_{NH}, ¹D_{CαH}, ¹D_{NC'}, and ²D_{HNC'}) and DNA (¹D_{HH}) were measured using a 1:1 complex of ¹⁵N/¹³C-labeled protein and unlabeled DNA. Dipolar couplings for the DNA (¹D_{NH} and ¹D_{CH}) and ¹D_{NH} dipolar couplings for the protein were measured on a 1:1 complex of ¹⁵N-labeled protein and ¹⁵N/¹³C-labeled DNA. ¹D_{NH} couplings were measured using a 2D IPAP (in-phase/anti-phase) ¹H-coupled {¹⁵N,¹H} HSQC experiment,⁶² ¹D_{NC'} and ²D_{HNC'} couplings were measured using a 2D ¹³C'-coupled/¹³C^α-decoupled(F₁) ¹H-¹⁵N HSQC experiment,⁶³ and ¹D_{CαH} couplings were measured from a 3D ¹H^α-coupled(F₁) HCA(CO)N experiment.⁶⁴ Other ¹D_{CH} couplings were measured from 2D ¹H-coupled (F₁) ¹H-¹³C HSQC and ¹H-coupled (F₂) ¹H-¹³C HSQC experiments. In addition, approximate ¹H-¹H dipolar coupling restraints for the DNA, grouped into ranges corresponding to strong, medium and weak intensity cross-peaks,⁶⁵ were derived from a 2D ¹²C-filtered COSY spectrum recorded on a sample of ¹⁵N/¹³C-labeled protein and unlabeled DNA in the bicelle liquid crystalline medium. The magnitudes of the axial (*D_a^{NH}*) and rhombic (*η*) components of the alignment tensor *D^{NH}* were obtained by examining the distribution of normalized residual dipolar couplings.⁶⁶ The value of *D_a^{NH}* is critically dependent on the concentration of bicelles which tends to vary from preparation to preparation. The values of *D_a^{NH}* ranged from 7.9 to 8.6 for both complexes. The rhombicity is independent of bicelle concentration, and had a value of 0.6 for the wild-type complex and 0.5 for the M9I complex.

Structure calculations

Approximate interproton distance and torsion angle restraints were derived from the NOE and coupling constant data as described by Omichinski *et al.*⁶⁷ Of the 26 phosphates groups, 14 had ³¹P shifts in the range –3.0 to –3.5 ppm, characteristic of a B-DNA.^{65,68} Broad torsion angle restraints for the DNA backbone of these nucleotides covering the ranges characteristic of both A and B-DNA were therefore employed to prevent problems associated with local mirror images: α = –70(±50)°, β = 180(±50)°, γ = 60(±35)°, ε = 180(±50)°, and ζ = –85(±50)°. The ³¹P shifts (wild-type/M9I complex) of five nucleotides, A12 (–3.01/–3.08 ppm), C25 (–3.42/–3.51 ppm), G22 (–5.13/–5.13 ppm), T21 (–5.37/–5.34 ppm) and A9 (–5.81/–5.79 ppm), lie outside the –3.0 to –3.5 ppm range, and the ³¹P shift of T20 for both the wild-type and M9I complex could not be observed. Consequently, the backbone α, β, γ, ε and ζ sugar-phosphate torsion angles associated with the phosphate groups of these

six nucleotides were not restrained. With the exception of the phosphate group of T20, strong $H4'(i)-^{31}P(i)$ correlations were observed in the $^1H-^{31}P$ spectra of the two complexes, indicating that the β torsion angles are in the *trans* conformation; no $H5'/H5''$ correlations, however, were observed, indicating a g^+ conformation for the γ torsion angles.⁶⁸ We also note that the measured $^3J_{H3-P}$ couplings (for 14 out of 26 phosphate groups) ranged from ≤ 1 to 7 Hz, consistent with ϵ torsion angles in the range 165–205°.⁶¹ In addition the δ (C5'-C4'-C3'-O3') torsion angle was restrained to 145(±20)° in those cases (24 out of 28) where a qualitative interpretation of the NOE, ROE and J coupling data indicated a sugar pucker in the C-2'-*endo*/C-1'-*exo* range, and to 75(±20)° in the three cases (A8, T19 and T20) where these data indicated a sugar pucker in the C-3'-*endo* range. No δ torsion angle restraint was employed for C14 since the NMB data were suggestive of a mixture of C-2' and C-3'-*endo* conformations. NOEs were grouped into four distance ranges: 1.8–2.7 Å (1.8–2.9 Å for NOEs involving NH protons), 1.8–3.3 Å (1.8–3.5 Å for NOEs involving NH protons), 1.8–5.0 Å, and 1.8–6.0 Å, corresponding to strong, medium, weak and very weak NOEs, respectively. In addition, 0.5 Å was added to the upper limit of interproton distance restraints involving methyl groups. Distances involving ambiguous NOEs, non-stereospecifically assigned methylene protons, methyl groups and the H^δ and H^ϵ protons of Tyr and Phe, were represented as a $(\sum r^{-6})^{-1/6}$ sum.⁶⁹

Structures were calculated by simulated annealing⁷⁰ using the NIH version (G.M.C., J. Kuszewski & C.D. Schweiters, available at <http://nmr.cit.nih.gov>) of XPLOR,⁷¹ which has been highly modified to incorporate numerous features relevant to NMR⁷² as well as new and highly efficient algorithms for torsion angle dynamics and minimization.⁷³ All simulated annealing⁷⁰ calculations were carried out in torsion angle space; the torsion angle dynamics algorithm employed a sixth-order predictor-corrector integrator with automatic time-step selection that varied during the course of the calculation.⁷³ The simulated annealing protocol employed was essentially that described by Omichinski *et al.*⁶⁷ with the difference that torsion angle dynamics rather than Cartesian coordinate dynamics were employed, and that the target function included a few additional terms. Bond lengths and angles were constrained to idealized covalent geometry.

The target function for simulated annealing comprised harmonic terms for covalent geometry (i.e. improper torsion angles used to define chirality and planarity, and bond lengths and angles associated with the closed ring systems of proline and the DNA sugar moieties; note that the other bonds and angles are held fixed by constraints); square-well potentials for the interproton distance, torsion angle, $^1H-^1H$ dipolar coupling⁷⁴ and hydrogen bonding restraints (excluding Watson-Crick hydrogen bonding restraints); harmonic potentials for the $^3J_{HN\alpha}$ couplings, $^{13}C^\alpha$ and $^{13}C^\beta$ secondary chemical shift, and heteronuclear dipolar coupling restraints;⁷⁵ harmonic potentials for the Watson-Crick hydrogen bonding distance restraints (six per base-pair;⁷⁶) and base-pair planarity restraints used to prevent undue buckling while allowing propeller twisting to occur;⁷⁶ and three terms for the non-bonded contacts. The latter comprise a quartic van der Waals repulsion term,⁷⁰ the torsion angle database potential term of mean force,⁷⁷ and the DELPHIC base-base positioning database potential of mean force.⁷⁸ No hydrogen bonding, electrostatic or 6-12 Lennard-Jones empirical potential energy terms were

present in the target function used for simulated annealing or restrained regularization.

The final force constants for the various terms in the target function were as follows: 1000 kcal mol⁻¹ Å⁻² for bond lengths; 500 kcal mol⁻¹ rad⁻² for angles and improper torsions; 30 kcal mol⁻¹ Å⁻² for the experimental distance restraints (interproton distances and hydrogen bonds); 200 kcal mol⁻¹ rad⁻² for torsion angle restraints; 1 kcal mol⁻¹ Hz⁻² for the $^3J_{HN\alpha}$ coupling constant restraints; 0.5 kcal mol⁻¹ ppm⁻² for the secondary ^{13}C chemical shift restraints; 1.0 kcal mol⁻¹ Hz⁻² for the $^1D_{NH}$, $^1D_{CH}$ and D_{HH} dipolar coupling restraints; 0.05 and 0.108 kcal mol⁻¹ Hz⁻² for the normalized (relative to $^1D_{NH}$) $^1D_{NC}$ and $^2D_{HNC}$ dipolar coupling restraints, respectively; 4 kcal mol⁻¹ Å⁻⁴ for the quartic van der Waals repulsion term (with the van der Waals radii set to 0.78 times their values used in the CHARMM PARAM19/20 protein and PARNAH1ER nucleic acid parameters); 1.0 for the DELPHIC torsion angle database potential; and 0.25 for the DELPHIC base-base positioning database potential.

The precision of the DNA coordinates in the present structures (~ 0.2 Å) is higher than that reported for previous protein-DNA structures from our group (0.3–0.5 Å) (for example, see references 67, 76, 80 and 81). This is due to three factors that are pertinent to the present structures relative to the previous ones: (a) the use of residual dipolar couplings for the DNA;^{65,76,78} (b) the higher (approximately double) density of intermolecular NOEs coupled with the higher precision of the backbone protein coordinates afforded by an extensive set of protein backbone residual dipolar couplings; and (c) the use of the DELPHIC base-base positional database potential.⁷⁸ The purpose of the base-base positional database potential, which in essence is a non-bonded contact term, is to bias sampling of base-base contacts (intra- and interstrand) during simulated annealing refinement to physically reasonable regions of conformational space within the range of possibilities that are consistent with the experimental NMR restraints.⁷⁹ The base-base positional database potential is based on 291 DNA crystal structures (including A, B, Z and unusual DNA structures, protein-DNA complexes, and drug-DNA complexes) solved at a resolution of 2 Å or better. The effects of the base-base positional database potential on the precision and accuracy of DNA coordinates obtained from NMR data has been discussed extensively by Kuszewski & Clore.⁷⁸ Using a combination of different approaches, including cross-validation against experimental NMR restraints, it was shown that the base-base positional database potential results in a large increase in accuracy (as well as a concomitant increase in precision) and obviates artifactual distortions in NMR-derived DNA structures arising from the limitations of conventional descriptions of the non-bonded contacts in terms of either Lennard-Jones van der Waals and electrostatic potentials or a simple van der Waals repulsion potential.⁷⁸ It was also shown by Kuszewski & Clore⁷⁸ that the base-base positional database potential does not in any way preclude unusual DNA structures and that in the case of the wild-type SRY-DNA complex the DNA between base-pairs 4–12, the region that displays significant distortions, is essentially identical for the structures calculated with and without the base-base positional database potential with an atomic rms difference of ~ 0.4 Å which is within the errors of the coordinates (cf. Figure 14 by Kuszewski & Clore⁷⁸). It should, however, be borne in mind that the precision of the coordinates is significantly greater than their accuracy, and the likely

accuracy of the present coordinates is probably somewhere in the range of $\sim 0.5\text{--}0.8 \text{ \AA}$.⁷⁸

The final ensemble of simulated annealing structures comprised 400 structures each for the wild-type and M9I complexes. The DNA bend angle was calculated with the program CURVES,⁸¹ and all other structural DNA parameters were calculated with the program CompDNA.⁸² Structure Figures were generated using the programs VMD-XPLOR⁸³ (available at <http://nmr.cit.nih.gov>), MOLMOL⁸⁴ and GRASP.⁸⁵

Protein Data Bank accession code

The coordinates have been deposited in the RCSB Protein Data Bank (accession codes 1J46 and 1J47 for the wild-type and M9I complexes, respectively).

Acknowledgements

We thank John Kuszewski, Dan Garrett, Charles Schwieters and Frank Delaglio for software support; Angela C. Murphy for amino acid analysis and Ettore Appella for mass spectrometry; and Ad Bax, Carole Bewley, Michael Caffrey, John Kuszewski, Charles Schwieters and Mary Starich for useful discussions. This work was supported by the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health (to G.M.C.).

© 2001 US Government

References

- Tjian, R. & Maniatis, T. (1994). Transcriptional activation: a complex puzzle with few easy pieces. *Cell*, **77**, 5-8.
- Grosschedl, R. (1995). Higher-order nucleoprotein complexes in transcription: analogies with site-specific recombination. *Curr. Opin. Cell. Biol.* **7**, 362-370.
- Bewley, C. A., Gronenborn, A. M. & Clore, G. M. (1998). Minor groove-binding architectural proteins: structure, function and DNA recognition. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 105-131.
- Bianchi, M. E. & Beltrame, M. (1998). Flexing DNA: HMG-box proteins and their partners. *Am. J. Hum. Genet.* **63**, 1573-1577.
- Maniatis, R., Falvo, J. V., Kim, T. H., Lin, C. H., Parekh, B. S. & Wathlet, M. G. (1998). Structure and function of the interferon- β enhanceosome. *Cold Spring Harbor Symp. Quant. Biol.* **63**, 609-620.
- Adhya, S., Geanakopoulos, M., Lewis, D. E., Roy, S. & Aki, T. (1998). Transcription regulation by repressosome and by RNA polymerase contact. *Cold Spring Harbor Symp. Quant. Biol.* **63**, 1-9.
- Murphy, F. V. I. V. & Churchill, M. A. (2000). Non-sequence-specific DNA recognition: a structural perspective. *Structure*, **8**, R83-R89.
- Weiss, M. A. (2001). Floppy SOX: mutual induced fit in HMG (high-mobility group) box-DNA recognition. *Mol. Endocrinol.* **15**, 353-362.
- Laudet, V., Stehelin, D. & Clevers, H. (1993). Ancestry and diversity of the HMG box superfamily. *Nucl. Acids Res.* **21**, 2493-2501.
- Grosschedl, R., Giese, K. & Pagel, J. (1993). HMG domain proteins: architectural elements in the assembly of nucleoprotein structures. *Trends. Genet.* **10**, 94-100.
- Bustin, M. & Reeves, R. (1996). High-mobility-group chromosomal proteins: architectural components that facilitate chromatin function. *Prog. Nucl. Acids Res. Mol. Biol.* **54**, 35-100.
- Werner, M. H., Huth, J. R., Gronenborn, A. M. & Clore, G. M. (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.
- Love, J. J., Li, X., Case, D. A., Giese, K., Grosschedl, R. & Wright, P. E. (1995). Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature*, **376**, 791-795.
- Murphy, F. V. I. V., Sweet, R. M. & Churchill, M. E. A. (1999). The structure of a chromosomal high mobility group protein-DNA complex reveals sequence-neutral mechanisms important for non-specific DNA recognition. *EMBO J.* **18**, 6610-6618.
- Allain, F. H.-T., Yen, Y.-M., Masse, J. E., Schultze, P., Dieckmann, T., Johnson, R. C. & Feigon, J. (1999). Solution structure of the HMG protein NHP6A and its interaction with DNA reveals the structural determinants for non-specific binding. *EMBO J.* **18**, 2563-2579.
- Ohndorf, U.-M., Rould, M. A., He, Q., Pabo, C. O. & Lippard, S. J. (1999). Basis for recognition of cisplatin-modified DNA by high-mobility-group proteins. *Nature*, **399**, 708-712.
- Goodfellow, P. N. & Lovell-Badge, R. (1993). SRY and sex determination in mammals. *Annu. Rev. Genet.* **27**, 71-92.
- Koopman, P. (1999). *Sry* and *Sox9*: mammalian testis-determining genes. *Cell. Mol. Life Sci.* **55**, 839-856.
- Parker, K. L., Schedl, A. & Schimmer, B. P. (1999). Gene interactions in gonadal development. *Annu. Rev. Physiol.* **61**, 417-433.
- O'Neill, M. J. & O'Neill, R. J. W. (1999). Whatever happened to SRY? *Cell. Mol. Life Sci.* **56**, 883-893.
- Haqq, C. M., King, C. Y., Donahoe, P. K. & Weiss, M. (1994). SRY recognizes conserved DNA sites in sex-specific promoters. *Proc. Natl Acad. Sci. USA*, **90**, 1097-1101.
- Haqq, C. M., King, C. Y., Ukiyama, E., Falsafi, S., Haqq, T. N., Donahoe, P. K. & Weiss, M. A. (1994). Molecular basis of mammalian sexual determination: activation of Müllerian inhibiting substance gene expression by SRY. *Science*, **266**, 1494-1500.
- Jeske, Y. W. A., Mishina, Y., Cohen, D., Behringer, R. & Koopman, P. (1996). Analysis of the role of *Amh* and *Fra-1* in the *Sry* regulatory pathway. *Mol. Reprod. Dev.* **44**, 153-158.
- Kent, J., Wheatley, S. C., Andrews, J. E., Sinclair, A. H. & Koopman, P. (1996). A male-specific role for *Sox9* in vertebrate sex determination. *Development*, **122**, 2813-2822.
- Morais da Silva, S., Jacker, A., Harley, V., Goodfellow, P. N., Swain, A. & Lovell-Badge, R. (1996). *Sox9* expression during gonadal development implies a conserved role for the gene in testis differentiation in mammals and birds. *Nature Genet.* **14**, 62-68.
- Sinclair, A. H., Berta, P., Palmer, M. S., Hawkins, J. R., Griffiths, B. L. & Smith, M. J. *et al.* (1990). A gene from the human sex determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature*, **346**, 240-244.

27. Südbeck, P. & Scherer, G. (1997). Two independent nuclear localization signals are present in the DNA-binding high-mobility group domains of SRY and SOX9. *J. Biol. Chem.* **272**, 27848-27852.
28. Harley, V. R., Lovell-Badge, R., Goodfellow, P. N. & Hextan, P. J. (1996). The HMG box of SRY is a calmodulin binding domain. *FEBS Letters*, **391**, 24-28.
29. Clore, G. M., Werner, M. H., Huth, J. R. & Gronenborn, A. M. (1997). Understanding 46X,Y sex reversal at the atomic level. In *Mechanisms of Transcription* (Ekstein, F. & Lilley, D. M. J., eds), pp. 265-273, Springer Verlag, Berlin.
30. Pontiggia, A., Rimini, R., Harley, V. R., Goodfellow, P. N., Lovell-Badge, R. & Bianchi, M. E. (1994). Sex-reversing mutations affect the architecture of SRY-DNA complexes. *EMBO J.* **13**, 6115-6124.
31. Clore, G. M. & Gronenborn, A. M. (1991). Structures of larger proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy. *Science*, **252**, 1390-1399.
32. Clore, G. M. & Gronenborn, A. M. (1998). Determining structures of larger proteins and protein complexes by NMR. *Trends Biotechnol.* **16**, 22-34.
33. Bax, A. & Grzesiek, S. (1993). Methodological advances in protein NMR. *Accts Chem. Res.* **26**, 131-138.
34. Tjandra, N. & Bax, A. (1997). Direct measurement of distances and angles in biomolecules by NMR in dilute liquid crystalline medium. *Science*, **278**, 1111-1114.
35. Tjandra, N., Omichinski, J. G., Gronenborn, A. M., Clore, G. M. & Bax, A. (1997). Use of dipolar ^1H - ^{15}N and ^1H - ^{13}C couplings in the structure determination of magnetically oriented macromolecules in solution. *Nature Struct. Biol.* **4**, 732-738.
36. Clore, G. M., Starich, M. R., Bewley, C. A., Cai, M. & Kuszewski, J. (1999). Impact of residual dipolar couplings on the accuracy of NMR structures determined from a minimal number of NOE restraints. *J. Am. Chem. Soc.* **121**, 6513-6514.
37. Clore, G. M. & Garrett, D. S. (1999). R-factor, free R and complete cross-validation for dipolar coupling refinement of NMR structures. *J. Am. Chem. Soc.* **121**, 9008-9012.
38. Garrett, D. S., Seok, Y.-J., Peterkofsky, A., Gronenborn, A. M. & Clore, G. M. (1999). Solution structure of the 40,000 M_r phosphoryl transfer complex between enzyme I and HPr. *Nature Struct. Biol.* **6**, 166-173.
39. Clore, G. M. (2000). Accurate and rapid docking of protein-protein complexes on the basis of intermolecular nuclear Overhauser enhancement data and dipolar couplings by rigid body minimization. *Proc. Natl Acad. Sci. USA*, **97**, 9021-9025.
40. Wang, G., Louis, J. M., Sondej, M., Seok, Y.-J., Peterkofsky, A. & Clore, G. M. (2000). Solution structure of the phosphoryl transfer complex between the signal transducing proteins HPr and IIA^{Glucose} of the *Escherichia coli* phosphoenolpyruvate:sugar phosphotransferase system. *EMBO J.* **19**, 5635-5649.
41. Mandel-Gutfreund, Y., Margalit, H., Jernigan, R. L. & Zhurkin, V. B. (1998). A role for CH—O interactions in protein-DNA recognition. *J. Mol. Biol.* **277**, 1129-1140.
42. Zhurkin, V. B., Ulyanov, N. B. & Ivanov, V. I. (1988). Mechanisms of DNA bending in the free state and in the nucleosome. In *DNA Bending and Curvature* (Olson, W. K., Sarma, M. H., Sarma, R. H. & Sundaralingam, M., eds), vol. 3, pp. 169-190, Adenine Press, New York.
43. Olson, W. K., Gorin, A. A., Lu, X.-J., Hock, L. M. & Zhurkin, V. B. (1998). DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl Acad. Sci. USA*, **95**, 11163-11168.
44. Olson, W. K. & Flory, P. J. (1972). Steric interactions in polyribonucleotides: a virtual bond model. *Biopolymers*, **11**, 1-23.
45. Selsing, E., Wells, R. D., Alden, C. J. & Arnott, S. (1979). Bent DNA: visualization of a base-paired and stacked A-B conformational junction. *J. Biol. Chem.* **254**, 5417-5422.
46. Keifer, J. R., Mao, C., Braman, J. C. & Beese, L. S. (1998). Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal. *Nature*, **391**, 304-307.
47. Geise, K., Amsterdam, A. & Grosschedl, R. (1991). DNA-binding properties of the HMG domain of the lymphoid-specific transcriptional regulator LEF-1. *Genes Dev.* **5**, 2567-2578.
48. Read, C. M., Cary, P. D., Preston, N. S., Lnenicek-Allen, M. & Crane-Robinson, C. (1994). The DNA sequence specificity of HMG boxes lies in the minor wing of the structure. *EMBO J.* **13**, 5639-5646.
49. McDowall, S., Argentaro, A., Ranganathan, S., Weller, P., Mertin, S. & Mansour, S. *et al.* (1999). Functional and structural studies of wild-type SOX9 and mutations causing campomelic dysplasia. *J. Biol. Chem.* **274**, 24023-24030.
50. Thompson, J. F. & Landy, A. (1988). Empirical estimation of protein-induced bending angles: applications to lambda site-specific recombination complexes. *Nucl. Acids Res.* **16**, 9687-9705.
51. Hardwidge, P. R., Den, R. B., Ross, E. D. & Maher, L. J., III (2000). Relating independent measures of DNA curvature: electrophoretic anomaly and cyclization efficiency. *J. Biomol. Struct. Dynam.* **18**, 219-229.
52. Cherny, D. I., Striker, G., Subramanian, V., Jett, S. D., Palecek, E. & Jovin, T. M. (1999). DNA bending due to specific p53 and p53 core domain-DNA interactions visualized by electron microscopy. *J. Mol. Biol.* **294**, 1015-1026.
53. Olson, W. K. & Zhurkin, V. B. (2000). Modeling DNA deformations. *Curr. Opin. Struct. Biol.* **10**, 286-297.
54. Louis, J. M., Martin, R. G., Clore, G. M. & Gronenborn, A. M. (1998). Preparation of uniformly isotope-labeled DNA oligonucleotides for NMR spectroscopy. *J. Biol. Chem.* **273**, 2374-2378.
55. Kim, J., Zwieb, C., Wu, C. & Adhya, S. (1989). Bending of DNA by gene-regulatory proteins: construction and use of a DNA bending vector. *Gene*, **85**, 15-23.
56. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, P. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX PIPES. *J. Biomol. NMR*, **6**, 277-293.
57. Garrett, D. S., Powers, R., Gronenborn, A. M. & Clore, G. M. (1991). A common sense approach to peak picking in two-, three- and four-dimensional spectra using automatic computer analysis of contour diagrams. *J. Magn. Reson.* **94**, 214-220.
58. Clore, G. M. & Gronenborn, A. M. (1989). Determination of three-dimensional structures of proteins and nucleic acids in solution by nuclear magnetic

- resonance spectroscopy. *CRC Crit. Rev. Biochem. Mol. Biol.* **24**, 479-564.
59. Bax, A., Vuister, G. W., Grzesiek, S., Delaglio, F., Wang, A. C., Tschudin, R. & Zhu, G. (1994). Measurement of homo- and heteronuclear J couplings from quantitative J correlation. *Methods Enzymol.* **239**, 79-106.
 60. Cornilescu, G., Delaglio, F. & Bax, A. (1999). Protein backbone angle restraints from searching a database for protein chemical shift and sequence homology. *J. Biomol. NMR*, **13**, 289-302.
 61. Clore, G. M., Murphy, E. C., Gronenborn, A. M. & Bax, A. (1998). Determination of three-bond $^1\text{H}3'-^{31}\text{P}$ couplings in nucleic acids and protein-nucleic acid complexes by quantitative J correlation spectroscopy. *J. Magn. Reson.* **134**, 164-167.
 62. Ottiger, M., Delaglio, F. & Bax, A. (1998). Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson.* **131**, 373-378.
 63. Delaglio, F., Torchia, D. A. & Bax, A. (1991). Measurement of ^{15}N - ^{13}C J couplings in staphylococcal nuclease. *J. Biomol. NMR*, **1**, 439-446.
 64. Bewley, C. A., Gustafson, K. R., Boyd, M. R., Covell, D. G., Bax, A., Clore, G. M. & Gronenborn, A. M. (1998). Solution structure of cyanovirin-N, a potent HIV-inactivating protein. *Nature Struct. Biol.* **5**, 571-578.
 65. Tjandra, N., Tate, S., Ono, A., Kainosho, M. & Bax, A. (2000). The NMR structure of a DNA dodecamer in an aqueous dilute liquid crystalline phase. *J. Am. Chem. Soc.* **122**, 6190-6200.
 66. Clore, G. M., Gronenborn, A. M. & Bax, A. (1998). A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. *J. Magn. Reson.* **133**, 216-221.
 67. Omichinski, J. G., Pedone, P. V., Felsenfeld, G., Gronenborn, A. M. & Clore, G. M. (1997). The solution structure of a specific GAGA factor-DNA complex reveals a modular binding mode. *Nature Struct. Biol.* **4**, 122-132.
 68. Gorenstein, D. G. (1994). Conformation and dynamics of DNA and protein-DNA complexes by P-31 NMR. *Chem. Rev.* **94**, 1315-1338.
 69. Nilges, M. (1993). A calculational strategy for the structure determination of symmetric dimers by ^1H NMR. *Proteins: Struct. Funct. Genet.* **17**, 297-309.
 70. Nilges, M., Gronenborn, A. M., Brünger, A. T. & Clore, G. M. (1988). Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints: application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. *Protein Eng.* **2**, 27-38.
 71. Brünger, A. T. (1993). *XPLOR: A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.
 72. Clore, G. M. & Gronenborn, A. M. (1998). New methods of structure refinement for macromolecular structure determination by NMR. *Proc. Natl Acad. Sci. USA*, **95**, 5891-5898.
 73. Schwieters, C. D. & Clore, G. M. (2001). Internal coordinates for molecular dynamics and minimization in structure determination and refinement. *J. Magn. Reson.* **In the press**.
 74. Tjandra, N., Marquardt, J. & Clore, G. M. (2000). Direct refinement against proton-proton dipolar couplings in NMR structure determination of macromolecules. *J. Magn. Reson.* **142**, 393-396.
 75. Clore, G. M., Gronenborn, A. M. & Tjandra, N. (1998). Direct refinement against residual dipolar couplings in the presence of rhombicity of unknown magnitude. *J. Magn. Reson.* **131**, 159-162.
 76. Huang, K., Louis, J. M., Donaldson, L., Lim, F.-L., Sharrocks, A. D. & Clore, G. M. (2000). Solution structure of the MEF2A-DNA complex: structural basis for the modulation of DNA bending and specificity by MADS-box transcription factors. *EMBO J.* **19**, 2615-2628.
 77. Kuszewski, J. & Clore, G. M. (2000). Source of and solutions to problems in the refinement of protein NMR structures against torsion angle potentials of mean force. *J. Magn. Reson.* **146**, 249-254.
 78. Kuszewski, J., Schwieters, C. D. & Clore, G. M. (2001). Improving the accuracy of NMR structures of DNA by means of a database potential of mean force describing base-base positional interactions. *J. Am. Chem. Soc.* **123**, 3903-3918.
 79. Starich, M. R., Wikström, M., Arst, H. N., Clore, G. M. & Gronenborn, A. M. (1998). The solution structure of a fungal AREA protein-DNA complex: an alternative binding mode for the basic carboxyl tail of GATA factors. *J. Mol. Biol.* **277**, 605-620.
 80. Starich, M. R., Wikström, M., Shumaker, S., Arst, H. N., Gronenborn, A. M. & Clore, G. M. (1998). The solution structure of the Leu22Æval mutant of the AREA DNA binding domain complexed with a TGATA core element defines a role for hydrophobic packing in the determination of specificity. *J. Mol. Biol.* **277**, 621-643.
 81. Lavery, R. & Sklenar, H. (1989). The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dynam.* **6**, 63-91.
 82. Gorin, A. A., Zhurkin, V. B. & Olson, W. K. (1995). B-DNA twisting correlates with base-pair morphology. *J. Mol. Biol.* **247**, 34-48.
 83. Schwieters, C. D. & Clore, G. M. (2001). The VMD-XPLOR visualization package for NMR structure refinement. *J. Magn. Reson.* **149**, 239-244.
 84. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51-55.
 85. Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights into interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281-296.
 86. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291.

Edited by P. E. Wright

(Received 16 May 2001; received in revised form 23 July 2001; accepted 1 August 2001)