Cell, Vol. 81, 705–714, June 2, 1995, Copyright © 1995 by Cell Press

Molecular Basis of Human 46X,Y Sex Reversal Revealed from the Three-Dimensional Solution Structure of the Human SRY–DNA Complex

Milton H. Werner,* Jeffrey R. Huth,*
Angela M. Gronenborn, and G. Marius Clore
Laboratory of Chemical Physics
National Institute of Diabetes and Digestive and Kidney Diseases
National Institutes of Health
Bethesda, Maryland 20892-0520

Summary

The solution structure of the specific complex between the high mobility group (HMG) domain of SRY (hSRY–HMG), the protein encoded by the human testis-determining gene, and its DNA target site in the promoter of the Müllerian inhibitory substance gene has been determined by multidimensional NMR spectroscopy. hSRY–HMG has a twisted L shape that presents a concave surface (made up of three helices and the N- and C-terminal strands) to the DNA for sequence-specific recognition. Binding of hSRY–HMG to its specific target site occurs exclusively in the minor groove and induces a large conformational change in the DNA. The DNA in the complex has an overall 70°–80° bend and is helically unwound relative to classical A- and B-DNA. The structure of the complex reveals the origin of sequence-specific binding within the HMG-1/HMG-2 family and provides a framework for understanding the effects of point mutations that cause 46X,Y sex reversal at the atomic level.

Introduction

Male sex determination in mammals is directed by the genetic information encoded on the Y chromosome, leading to specialization of the embryonic gonads into testes (Goodfellow and Lovell-Badge, 1993; Haqq et al., 1994). A key component is the protein encoded by the human testis-determining gene *SRY* (for sex-determining region Y), mutations in which are responsible for 15% of male to female sex reversal, that is, 46X,Y females (Goodfellow and Lovell-Badge, 1993; Gustafson and Donahoe, 1994). The SRY protein is a transcriptional activator of the Müllerian inhibiting substance (*MIS*) gene, whose product is responsible for the regression of the female Müllerian ducts (the precursor of the uterus, fallopian tubes, and upper vagina) in male embryos. In addition, SRY may act as a repressor of the P450 aromatase gene, whose product is responsible for the conversion of testosterone to estradiol during female gonadal development (Haqq et al., 1993).

The human *SRY* (*hSRY*) gene codes for a protein of 203 residues that comprises three domains: an N-terminal domain, a central 77 residue DNA-binding domain consisting of a single high mobility group (HMG) box, and a C-terminal domain (Sinclair et al., 1990). Comparison of the sequences of SRY from different species indicates that only the DNA-binding HMG box domain is conserved (Whitfield et al., 1993; Tucker and Lundrigan, 1993). The HMG box is an ~80 residue domain that mediates the minor groove DNA binding of a large family of eukaryotic proteins known as the HMG-1/HMG-2 family (Laudet et al., 1993). This family can be broadly divided into two functional subclasses. The first subclass comprises transcription factors that bind sequence specifically to DNA, such as SRY, SOX5, lymphoid enhancer factor 1 (LEF-1), and T cell factor 1 (TCF-1). The second subclass shows little or no specificity and is typified by HMG-1 and HMG-2, which are abundant in chromatin, and upstream binding factor, which is a general transcription factor.

To establish the atomic and molecular basis of SRY-dependent 46X,Y sex reversal, we have determined, by means of multidimensional nuclear magnetic resonance (NMR), the solution structure of a specific complex of the DNA-binding domain of hSRY (hereafter referred to as hSRY–HMG) with a DNA octamer comprising its specific target site in the *MIS* promoter. In addition, this structure provides a basis for understanding the general structural determinants of minor groove binding by HMG domains.

Results

Structure Determination

The structure of the hSRY–HMG–DNA complex was solved by means of multidimensional double and triple resonance heteronuclear-filtered and heteronuclear-edited NMR spectroscopy (Clore and Gronenborn, 1994), making use of the DNA octamer at natural isotope abundance (that is, ^{12}C and ^{14}N) and uniformly (>95%) ^{15}N- and $^{15}N/^{13}C$-labeled hSRY–HMG. The sequence of the octamer, (5′-dGCACAAAC)·(5′-dGTTTGTGC), was obtained by comparison of the DNase footprint of intact hSRY on *MIS* (Haqq et al., 1993) to the binding site consensus sequence previously identified for hSRY, as well as the related proteins TCF-1 and LEF-1 (van de Wetering and Clevers, 1992; Giese et al., 1992; Harley et al., 1994). The structure calculations, using simulated annealing, were based on a total of 1805 experimental NMR restraints. A summary of the structural statistics is provided in Table 1, and a best fit superposition of the final 35 simulated annealing structures is shown in Figure 1. Residues 1–3 and 75–78 at the N- and C-termini, respectively, are disordered in solution. Excluding these residues, the atomic root-mean-square (rms) distribution of the 35 simulated annealing structures about their mean coordinate positions (that is, the precision of the coordinates) is 0.59 ± 0.05 Å for the protein backbone (N, Cα, C, and O atoms) plus the DNA and 0.96 ± 0.06 Å for all protein atoms plus DNA.

*The first two authors contributed equally to this work.

Table 1. Statistics

Statistics	<SA>	$\overline{(SA)}r$
Rms deviations from experimental distance restrants (Å)[a]		
All (1379)	0.040 ± 0.003	0.038
Protein		
Interresidue sequential (|i − j| = 1) (290)	0.027 ± 0.008	0.020
Interresidue short range (1 < |i − j| ≤5) (221)	0.034 ± 0.005	0.025
Interresidue long range (|i − j|) > 5) (107)	0.033 ± 0.008	0.024
Intraresidue (238)	0.017 ± 0.008	0.013
H bonds (70)	0.055 ± 0.009	0.053
DNA		
Intraresidue (206)	0.032 ± 0.002	0.027
Sequential intrastrand (96)	0.048 ± 0.008	0.057
Interstrand (36)	0.052 ± 0.004	0.050
H bonds (40)	0.123 ± 0.004	0.123
Protein–DNA (75)	0.054 ± 0.011	0.054
Rms deviations from experimental dihedral restraints (°) (225)[a]	0.67 ± 0.07	0.77
Rms deviations from experimental $^3J_{HN\alpha}$ coupling constants (Hz) (56)[a]	0.58 ± 0.03	0.57
Rms deviations from experimental ^{13}C shifts		
$^{13}C\alpha$ (ppm) (73)	1.02 ± 0.03	1.03
$^{13}C\beta$ (ppm) (72)	1.01 ± 0.02	1.12
Deviations from idealized covalent geometry		
Bonds (Å) (1849)	0.006 ± 0.0002	0.007
Angles (°) (3331)	1.352 ± 0.009	1.390
Impropers (°) (987)	0.528 ± 0.046	0.675
E_{L-J} (kcal/mol)[b]	−540 ± 8	−524

The notation of the NMR structures is as follows: <SA> are the final 35 simulated annealing structures; $\overline{SA}$ is the mean structure obtained by averaging the coordinates of the individual SA structures best fitted to each other (with respect to residues 4–74 of the protein and base pairs 1–8 of the DNA); $\overline{(SA)}r$ is the restrained minimized mean structure obtained by restrained regularization of the mean structure $\overline{SA}$. The number of terms for the various restraints is given in parentheses.

[a] None of the structures exhibited distance violations greater than 0.5 Å, dihedral angle violations greater than 5°, or $^3J_{HN\alpha}$ coupling constant violations greater than 2 Hz.

[b] E_{L-J} is the Lennard–Jones van der Waals energy calculated with the CHARMM PARAM19/20 protein and PARNAH1ER1 DNA parameters (Brooks et al., 1983) and is not included in the target function for simulated annealing or restrained minimization.

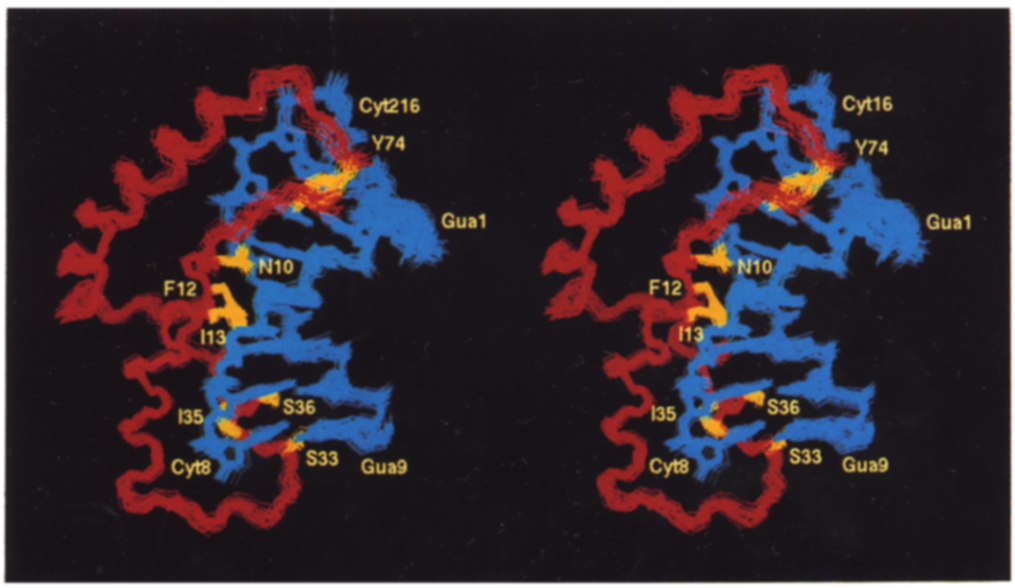

Figure 1. Stereoview Showing a Best Fit Superposition of the 35 Simulated Annealing Structures of the Specific Complex of hSRY–HMG with DNA

The backbone (N, Cα, and C) atoms (residues 4–74) of hSRY–HMG are shown in red, side chains that contact the DNA bases in yellow, and all nonhydrogen atoms of the DNA in blue.

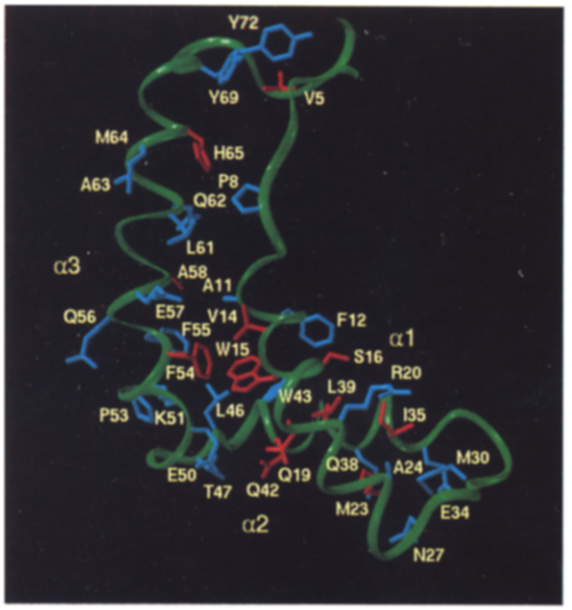

Figure 2. Ribbon Drawing of hSRY–HMG

The backbone ribbon is shown in green, and side chains with less than 20% and between 20% and 50% of their surface (relative to an extended Gly-X-Gly tripeptide segment) accessible to solvent are shown in red and blue, respectively.

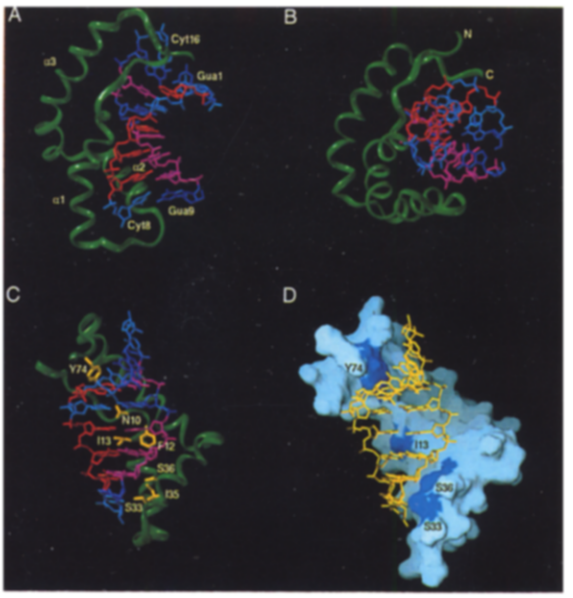

Figure 3. The Interaction of hSRY–HMG with DNA

Three views (A–C) are displayed. The protein is shown as a schematic ribbon drawing in green, and the color coding used for the DNA bases is red for A, lilac for T, dark blue for G, and light blue for C. Side chains that contact the DNA bases are depicted in yellow in (C). (D) shows the same view as in (C) with the molecular surface of the protein shown in gray and the DNA atoms in yellow. The patches of blue on the protein surface indicate the location of the side chains of four of the seven residues that interact with the DNA bases.

Tertiary Structure of the hSRY–HMG Domain

hSRY–HMG has a twisted letter L or boomerang shape, with the long and short arms ~28 Å and ~22 Å in length, respectively (Figure 2). The structure is composed of irregular N-terminal (residues 3–9) and C-terminal (residues 71–74) strands that lie directly opposite each other, although they are not hydrogen bonded, and three helices (residues 10–26, 33–45, and 47–68). Helix 1 has a slightly bent appearance, owing to a distortion at Asp-18. Helix 3 exhibits an ~50° kink at the junction of Pro-53 and Phe-54. The long arm of the L is formed by helix 3 and the N-terminal strand (residues 3–9), while the short arm of the L is formed by helices 1 and 2. Helix 1 is oriented at ~120° to helix 2 and ~130°–140° to helix 3. Helices 2 and 3 are approximately orthogonal to each other. The orientation of the long and short arms is maintained by the packing of Ala-11, Phe-12, Val-14, Trp-15, and Gln-19 of helix 1; Leu-39, Trp-43, and Leu-46 of helix 2; and Phe-54, Phe-55, and Ala-58 of helix 3. The aromatic rings of Trp-15 and Trp-43 are stacked upon one another and are arranged approximately orthogonally to the aromatic rings of Phe-12 and Phe-54. The top of the long arm is held in place by the packing of Val-5, Tyr-69, and Tyr-72 at the N- and C-termini, while the middle of the long arm is stabilized by the packing of Pro-8, Leu-61, Gln-62, and His-65. The loop at the end of the short arm that connects helices 1 and 2 is stabilized by the packing of Arg-20, Met-23, Met-30, Asn-27, and Ile-35. There are two potential stabilizing side chain electrostatic interactions between the backbone carbonyl of Asp-3 and the hydroxyl of Tyr-72 at the N- and C-termini and between the carboxyamide of Gln-19 and the carboxylate of Glu-50 in helices 1 and 3.

With the exception of the hydrophobic cluster at the N- and C-termini formed by Val-5, Tyr-69, and Tyr-72, the overall structure of hSRY–HMG is similar to that of the previously determined NMR structures of two nonspecific DNA-binding domains of the HMG-1/HMG-2 family (rat HMG-1 B and Drosophila HMG-D), both of which have 20%–25% sequence identity with hSRY–HMG (Weir et al., 1993; Read et al., 1993; Jones et al., 1994). The regions corresponding to the N- and C-terminal strands of hSRY–HMG, however, are disordered in both HMG-1 B and HMG-D. This is probably due to the replacement of Val-5 in hSRY–HMG by Pro in HMG-1 B and HMG-D, resulting in substantial structural instability at the N- and C-termini. A best fit superposition of the coordinates of hSRY–HMG and HMG-1 B indicates that the Cα atoms of 55 residues, comprising the three helices, can be superimposed with an atomic rms difference of 2 Å. Comparison of hSRY–HMG and HMG-D, on the other hand, reveals that the Cα atoms of only 36 residues, comprising helix 1, helix 2, and the middle portion of helix 3, can be superimposed within an rms of 2 Å.

Structure of the DNA Target Site

Upon binding to hSRY–HMG, the DNA undergoes a profound structural change from B-type DNA in the free state (as evidenced by NMR) to an underwound form that has features intermediate between A- and B-DNA (Figures 1, 3, and 4). The rms difference between the bound DNA octamer in the complex and classical A-DNA is 2.4 Å, com-

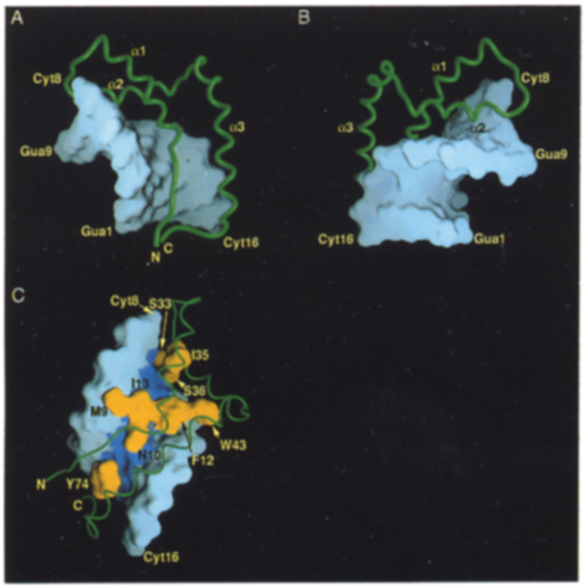

Figure 4. Complex of hSRY–HMG with DNA Illustrating the Conformation of the DNA

(A and B) Views in the minor (A) and major (B) grooves of the DNA illustrating the bend in the DNA, the widening of the minor groove, and the compression of the major groove, with the molecular surface of the DNA shown in gray and a ribbon drawing of the protein backbone in green.

(C) View of the complex in the minor groove showing the molecular surface of the DNA backbone (gray) and bases (blue), a ribbon drawing of the hSRY–HMG backbone (green), and the molecular surface of selected side chain atoms (yellow), illustrating the T-shaped wedge that contacts base pairs 4–6 (Met-9, Asn-10, Phe-12, Ile-13, and Trp-43) and the anchor points that contact the two ends of the DNA octamer (Tyr-74 at base pair 3; Ser-33, Ile-35, and Ser-36 at base pairs 7 and 8).

pared with a value of 4.2 Å for that between the bound DNA and classical B-DNA and between classical A- and B-DNA (Saenger, 1984). The DNA helix is severely underwound with an average local interbase pair helical twist of 26° ± 6°, ranging from a low value of ~19° between base pairs 1 and 2 and base pairs 7 and 8 to a high value of ~34° between base pairs 4 and 5. This compares with a local helical twist of 36° for B-DNA and 31° for A-DNA. As a result, the minor groove is shallow and significantly expanded, with a width of 9.4 ± 0.6 Å compared with 4.0 Å in B-DNA and 6.2 Å in A-DNA. Concomitantly, the major groove is substantially compressed (Figure 4B). The sugar puckers range from O1′-endo to C3′-endo for 14 of 16 deoxyribose rings, typical of A-DNA, with the exception of A5 and G9, which have C1′-exo to C2′-endo conformations, typical of B-DNA. The local interbase pair rise has an average value of 4.1 ± 0.3 Å, which is larger than that for either B-DNA (3.6 Å) or A-DNA (3.4 Å). Finally, the sequential phosphate–phosphate distances have an average value of 6.3 ± 0.4 Å, which is closer to that of B-DNA (6.5 Å) than A-DNA (5.6 Å). This distance, however, is significantly reduced to a value of 5.4 Å between base pairs 5 and 6, which is the site of partial intercalation by Ile-13.

The DNA in the complex is bent by ~70°–80° in the direction of the major groove (Figures 3 and 4), in agreement with the results obtained from electrophoretic mobility circular permutation shift assays (Ferrari et al., 1992). The bend is principally accomplished by large positive local interbase pair roll angles for 6 of the 7 base steps (ranging in value from 10.5° between base pairs 7 and 8 up to 35° between base pairs 2 and 3), in conjunction with the helical unwinding described above. This contrasts with values of 10.8° and −3.6° for the local interbase pair roll angles in classical A- and B-DNA, respectively (Saenger, 1984).

Comparison with the structure of the DNA in the complex with TATA-binding protein (TBP) (Kim et al., 1993a, 1993b), which also binds in the minor groove, reveals numerous features in common. These include the A-like sugar puckers, except at the sites of intercalation, where they are B-like; the large positive roll angles; the helix unwinding and concomitant widening of the minor groove; the increase in local helix rise relative to A- and B-DNA; and the B-like local slide. There are some interesting differences, however. Thus, the helix is ~20% more underwound and the roll angles are on average ~70% larger in the TBP complex than in the hSRY–HMG complex, while the minor groove is ~10% wider in the hSRY–HMG complex than in the TBP complex.

The Overall Structure of the Complex

The DNA is located in the concave surface of the L-shaped hSRY–HMG domain, and binding occurs exclusively in the minor groove of the DNA, in agreement with methyl and diethylpyrocarbonate interference and inosine replacement experiments (van de Wetering and Clevers, 1992). In the view shown in Figures 3C and 3D, it can be seen that the binding surface is formed by helices 1 and 3, bounded at the bottom by a ridge comprising helix 2 and at the top by a ridge comprising the N- and C-terminal strands. In contrast, the recognition of the minor groove by TBP is provided by the concave surface of a 10-stranded antiparallel β sheet (Kim et al., 1993a, 1993b). Thus, the structural scaffold of the DNA-binding surface is entirely different in hSRY–HMG and TBP. The conformation of the DNA, which is severely distorted from that of classical B-DNA, follows the contours of the concave binding surface perfectly (Figures 3 and 4). Hence, this interaction represents a classical example of induced fit. The orientation of the DNA with respect to the protein can be determined unambiguously on the basis of a qualitative interpretation of the nuclear Overhauser effects (NOEs) from Ser-33, Ile-35, and Ser-36 to the sugars of C8, G9, and T10; of the NOEs from the side chain of Ile-13 to the imino protons of T11 and T12, the H2 protons of A5 and A6, and the sugar protons of A5 and A6; and of the NOEs from the aromatic ring of Tyr-74 to the H2 proton of A3 and the sugar of C4.

Upon binding to DNA, there is a reduction of 1050 Å^2 in the accessible surface area of hSRY–HMG (corresponding to a 16% decrease) and a decrease in the calculated solvation free energy of folding of ~6.4 kcal/mol (Eisenberg and McLaghlan, 1986). This hydrophobic effect can clearly make a sizeable contribution to the observed equilibrium dissociation constant (~10 nM; Pontiggia et al., 1994). The two principal areas of bending occur between base pairs 5 and 6 as a result of the partial intercalation

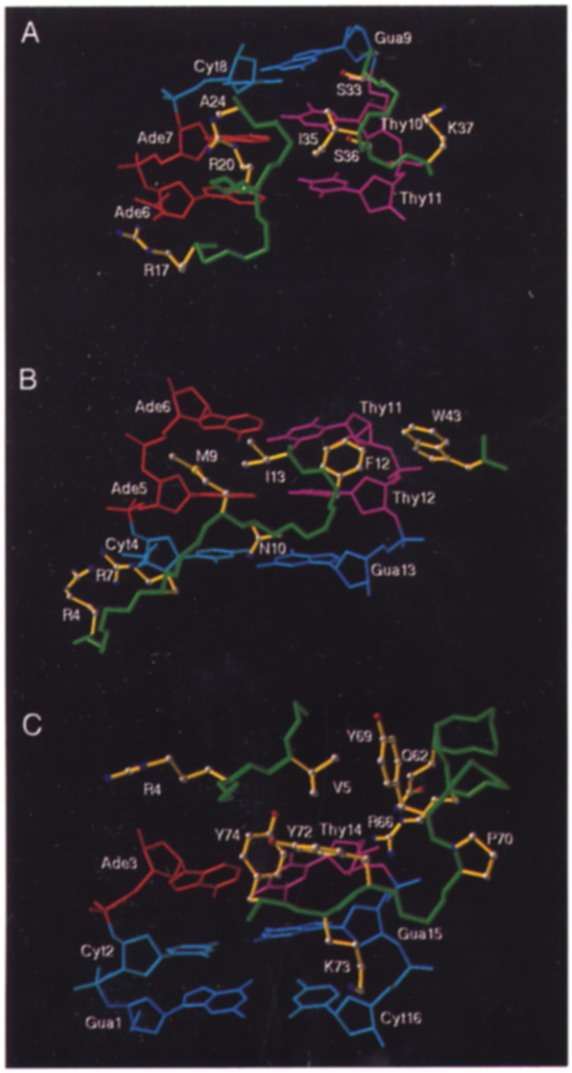

Figure 5. Side Chain Interactions between hSRY–HMG and the DNA
All contacts are in the minor groove. The polypeptide backbone is shown in green, the side chain bonds in yellow, and the side chain carbon, oxygen, and nitrogen atoms in gray, red, and blue, respectively. The color coding for the DNA is the same as in Figure 4.

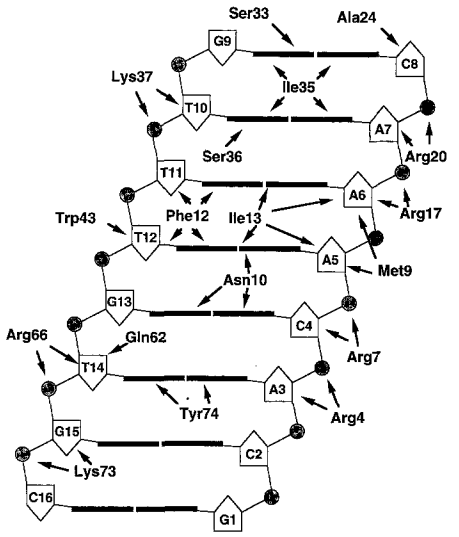

Figure 6. Summary of the Contacts between hSRY–HMG and DNA
The DNA is represented as a cylindrical projection viewed in the minor groove with the bases depicted as thick black lines, the deoxyribose sugar rings as pentagons, and the phosphates as stippled circles.

of Ile-13, as predicted by King and Weiss (1993), and between base pairs 2 and 3, as the DNA is pushed away from the body of the protein by the ridge formed by Lys-73 and Tyr-74 (see Figures 3C and 3D). Widening of the minor groove appears to be mediated by five residues that form a T-shaped wedge in direct contact with the central base pairs of the DNA octamer (Figure 4C). Met-9, Phe-12, Ile-13, and Trp-43 form a hydrophobic wedge across base pairs 5 and 6, anchored to base pairs 4 and 5 by electrostatic interactions involving Asn-10. The residues constituting the central portion of the wedge (Phe-12 and Ile-13) and the stem of the T (Asn-10) bind to the DNA bases, while the residues at the wings of the wedge (Met-9 and Trp-43) bind to the DNA sugar–phosphate backbone and pry open the minor groove (Figure 4C). In addition, hSRY–HMG is anchored at the two ends of the DNA by Tyr-74 at base pair 3 and by Ser-33, Ile-35, and Ser-36 at base pairs 7 and 8 (Figure 4C).

Contacts between hSRY–HMG and DNA

Seven residues (Asn-10, Phe-12, Ile-13, Ser-33, Ile-35, Ser-36, and Tyr-74) contact the DNA bases (Figures 5 and 6). The hydroxyl groups of Ser-33 and Ser-36 hydrogen bond to the N2 atom of G9 and the O2 of T10, respectively (Figure 5A), while the carboxyamide group of Asn-10 is involved in electrostatic interactions with the N2 atom of G13, the O2 atom of C4, and the N3 atom of A5 (Figure 5B). The two aromatic rings, Phe-12 and Tyr-74, are oriented orthogonally to the bases, a structural feature found in numerous examples of aromatic–aromatic interactions in proteins. Specifically, Phe-12 is packed against the bases T11 and T12 (Figure 5B), while Tyr-74 is packed against the bases of A3 and T14 (Figure 5C). The latter is responsible for the large roll angle between base pairs 2 and 3. In addition, the hydroxyl group of Tyr-74 may form a hydrogen bond with the O2 atom of T14. Ile-13 is partially intercalated between A5 and A6 (Figure 5B), thereby resulting in a large local roll angle at this base pair step. Finally, Ile-35 is packed against base pairs 7 and 8 (Figure 5A), but, unlike Ile-13, does not intercalate between the base pairs.

In addition to interactions with the bases, there are numerous other electrostatic and hydrophobic interactions involving the phosphates and sugars, respectively (Figures 5 and 6). Specifically, Arg-4, Arg-7, Arg-17, and Arg-20 form salt bridges to the phosphates of C4, A5, A7, and C8, respectively, on the sense strand of the DNA, while Lys-37, Arg-66, and Lys-73 form salt bridges to the phosphates of T11, T15, and T16, respectively, on the antisense strand. Finally, the remaining contacts are hydrophobic, involving 11 of the 16 sugars and the aliphatic

portions of several Arg side chains (Arg-4, Arg-7, Arg-17, Arg-20, and Arg-66), Met-9, Phe-12, Ala-24, and Trp-43. These electrostatic and hydrophobic interactions comprise a framework that stabilizes the bent DNA and act as a scaffold upon which base-specific interactions are imposed.

Discussion

Implications for DNA Binding Specificity of HMG Domains

A comparison of the amino acid sequences from a variety of specific and nonspecific HMG domains (Laudet et al., 1993) indicates that, in general, residues that interact with the sugar–phosphate backbone tend to be either conserved or substituted conservatively. The distinction between the two subclasses is found in residues that contact the bases and in certain core-packing amino acids.

In the case of the specific DNA-binding domains, the packing of Val-5, Tyr-69, and Tyr-72 at the N- and C-termini appear to be critical for the correct presentation of the DNA-binding surface made up of the side chains of Arg-4, Lys-73, and Tyr-74 (Figures 2 and 5C). The sequence-specific HMG domains have either Val or Ile at position 5. Mutation of Val-5 in hSRY by the longer Leu side chain leads to a substantial reduction in specific DNA binding affinity (Harley et al., 1992). The nonspecific HMG domains predominantly have the shorter Pro side chain at position 5. Thus, it seems mandatory that a β-branched aliphatic side chain is necessary for stable and optimal packing at this position and that any changes are sufficient to perturb specific DNA binding. The importance of the N- and C-termini is further supported by the results on a chimeric HMG domain composed of the N- and C-termini of the specific TCF-1 HMG domain and helices 1, 2, and a portion of helix 3 of the nonspecific HMG-1 B domain that retained the ability to both bend and bind DNA specifically (Read et al., 1994). Similar arguments hold for the stabilization of the twisted L shape by the hydrohobic core between the stem and base of the HMG domain. A key residue in this regard is Phe-55 in hSRY–HMG (Figure 2), which is either retained or substituted conservatively by a Tyr or Val in the specific HMG domains. In the nonspecific domains, on the other hand, Phe-55 is most often substituted by a Glu, perhaps destabilizing the hydrophobic core and thereby leading to less efficient unwinding of the DNA target. Indeed, the twist between the stem and base of the L reinforces the helical unwinding induced by the partial intercalation of Ile-13 by presenting the proper complementary surface to the DNA. Hence, destabilization of the hydrophobic core could, in part, disrupt the surface complementarity, leading to a more weakly bound protein and a concomitant loss of DNA sequence discrimination.

In addition, the nonspecific HMG domains lack a number of critical residues that make base-specific contacts in the structure of the hSRY–HMG complex. Asn-10 is conserved in all the specific HMG domains and is involved in hydrogen bonding or electrostatic interactions with the bases of C4, A5, and G13 (Figure 5B). In the nonspecific HMG domains, Asn-10 is substituted by Ser, Thr, or Arg. The side chains of Ser and Thr would be too short to make appropriate contact, while the longer Arg side chain is likely to introduce steric clash. A similar argument can be made for Ser-36, which hydrogen bonds with the base of T10 (Figure 5A). In the specific HMG domains, position 36 is occupied by Ser or Asn. In the nonspecific HMG domains, position 36 tends to be occupied by Ala, Val, or Lys. The impact of this change, however, is likely to be weaker than that for Asn-10 as a few nonspecific domains retain Ser or Thr at this position.

The structure of the complex of hSRY–HMG with DNA also provides an explanation for the different sequence specificities of the SOX subfamily (which includes hSRY, mouse SRY, and the SOX cluster of HMG domains) versus the LEF/TCF-1 subfamily. The DNA target sites for the SOX subfamily have a strong preference for A at the position of base 3, while those for the LEF/TCF-1 family display a strong preference for T at bases 2 and 3 (Giese et al., 1992). The main amino acid sequence difference between the two subfamilies is located at positions 73 and 74, which are Lys and Tyr in the SOX subfamily and Ser and Ala in the LEF/TCF-1 subfamily. In the hSRY–HMG complex, Lys-73 forms a phosphate contact to C16, while Tyr-74 is packed orthogonally to the bases of A3 and T14 and may form a hydrogen bond with the O2 atom of T14 (Figure 5C). The latter is precluded when T14 is replaced by A in the LEF/TCF-1 target site. Conversely, substitution of Tyr-74 by Ala in the LEF/TCF-1 subfamily removes a bulky hydrophobic group capable of extensive interactions with the bases. The substitution of Lys-73 by a Ser in LEF/TCF-1, however, permits an alternative hydrogen bond between the hydroxyl group of Ser-73 and the O2 atom of T2 (or C2 in the *MIS* promoter sequence). This interpretation is reinforced by the observation that LEF-1 and TCF-1 can bind with significant affinity to SRY sequences, while the reverse is not true (Giese et al., 1992).

The Molecular Basis of Human 46X,Y Sex Reversal and the Control of Transcription

The amino acid sequences of SRY derived from different organisms diverge considerably outside the HMG domain (Whitfield et al., 1993; Tucker and Lundrigan, 1993). Further, all clinical mutations in hSRY resulting in phenotypic 46X,Y sex reversal, with the exception of only a single nonsense mutation (Tajima et al., 1994), have been found to occur exclusively in the DNA-binding domain (Goodfellow and Lovell-Badge, 1993; Berta et al., 1990; Hawkins et al., 1992a, 1992b; McElreavy et al., 1992; Affara et al., 1993; Braun et al., 1993). These observations have led to the hypothesis that the primary influence of SRY on transcriptional regulation (and hence on gonadal differentiation) is largely a consequence of the structural effects induced by the protein at specific promoter targets in vivo. The structure of the specific complex of hSRY–HMG with DNA therefore provides a framework for explaining the effects of clinical mutations at the atomic level.

From the genetic standpoint, naturally occurring point mutations in hSRY are of two types: inherited mutations with variable penetrance and de novo mutations with full penetrance (Goodfellow and Lovell-Badge, 1993). Clearly,

the functional effects of the former must be considerably less severe than those of the latter. From a structural view point, the point mutations also fall into two categories: those that affect the packing of residues within the protein core and those that involve residues that directly contact the DNA.

Three inherited point mutations, Val-5→Leu, Ile-35→Met, and Phe-54→Ser, result in packing defects that would be expected to destabilize the protein. The first mutant reduces DNA binding significantly, the second either slightly or not at all, and the third has no effect (Harley et al., 1992; Pontiggia et al., 1994). Val-5 is crucial for the correct packing of Tyr-69 and Tyr-72 (Figure 2) that in turn serves to position Lys-73 and Tyr-74 appropriately for interaction with the DNA (Figure 5C). A longer Leu side chain will distort this region of the protein. The branched side chain of Ile-35 stabilizes the interaction between helix 1 and 2 by packing against Arg-20 and Met-23 (Figure 2) and, in addition, interacts with the bases of base pairs 7 and 8 (Figure 5A). Substitution by the unbranched Met side chain will destabilize some of these interactions. Phe-54 is a buried residue within the hydrophobic core that packs against Ala-11, Val-14, Trp-15, and Phe-55 at the junction of the base and stem of hSRY–HMG (Figure 2). Substitution by the polar Ser at this position would be anticipated to destabilize the protein. This effect, however, may not be sufficient to perturb the DNA binding properties of the Phe-54→Ser mutant, but may accelerate its degradation within the cell. The latter may also occur with the other two inherited mutations. Indeed, if increased degradation as a result of decreases in protein stability were the primary effect of the familial mutations, this would provide a simple explanation for their variable penetrance.

There are five other packing defect point mutants, four of which are de novo (Met-23→Thr, Leu-46→His, Ala-58→Thr, and Tyr-72→Cys), with the fifth (Lys-51→Ile) of unknown origin as relatives are unavailable. DNA binding data are only available for the Lys-51→Ile mutation, which displays reduced DNA binding affinity in the isolated HMG domain (Pontiggia et al., 1994) and no specific DNA binding in the full-length protein (Harley et al., 1992). This suggests that Lys-51 may play a role in the interaction of the HMG domain with the rest of the protein. The aliphatic portion of the side chain of Lys-51 is packed between Leu-46 and Phe-55 (Figure 2). Introduction of a branched Ile side chain with two bulky methyl groups at this position may disrupt Phe-55 from the hydrophobic core and partially displace Leu-46, thereby destabilizing the packing of helices 2 and 3. The Met-23→Thr mutant will destabilize the loop between helices 1 and 2, as a long aliphatic side chain would no longer be available to pack against Asn-27 and Met-30 (Figure 2). Leu-46 is packed between Lys-51, Trp-15, and Trp-43 so that the introduction of a polar His residue in the Leu-46→His mutant will destabilize the packing of helices 1 and 3. A similar effect will occur in the Ala-58→Thr mutation. Ala-58 is located in helix 3 and packed against Ala-11 and Val-14 in helix 1 (Figure 2). The introduction of the longer Thr side chain at position 58 will therefore introduce a steric clash between helices 1 and 3. Finally, the Tyr-72→Cys mutant, like the Val-5→Leu mutant, will disrupt the conformation of the N- and C-terminal strands by removing a key hydrophobic residue in the Val-5, Tyr-69, and Tyr-72 hydrophobic triad, thereby affecting the interaction of Lys-73 and Tyr-74 with the DNA. The increased penetrance of the Tyr-72→Cys mutation, relative to the Val-5→Leu mutation, can be attributed to two factors: first, the less conservative nature of the Tyr-72 substitution, which would therefore be expected to have a greater impact on the conformational stability of the N- and C-terminal strands; second, the possibility that the Tyr-72→Cys mutation may result in the formation of an undesired intermolecular disulfide bond.

The remaining four de novo point mutations identified to date involve residues that contact the DNA. The Arg-7→Gly mutant removes hydrophobic contacts with the sugar of C4 and a salt bridge to the phosphate of A5 (Figure 5B). Owing to shortening of the side chain, the Met-9→Ile mutant will disrupt extensive van der Waals contacts with the deoxyribose of A5 and A6 and with the O3′ atom of A6 that is located approximately at the hinge point of the bend between base pairs 5 and 6 (Figure 5B). Interestingly, this mutation reduces both the strength of binding and the extent of DNA bending (Pontiggia et al., 1994). The Gly-40→Arg mutation is likely to introduce a steric clash with the sugar–phosphate backbone of T12 (Figures 5 and 6) and with Trp-43 (Figure 2), thereby displacing a critical residue of the wedge that drives the helical unwinding of the DNA (Figure 4C). The conservation of Gly at position 40 in the sequence-specific HMG domains supports the notion that the local packing in this region of the interface is critical for the formation of a stable complex. Finally, the Ile-13→Thr mutation would be expected to have one of the most profound effects on DNA binding of all the mutants as the short polar Thr would not be able partially to intercalate between base pairs 5 and 6 (Figure 5B), thereby removing one of the principal determinants of bending and helical unwinding in the complex. Indeed, the affinity of this mutant for DNA is reduced by almost two orders of magnitude relative to that of wild type, and the mutant complex has a lifetime of less than 5 ms, compared with greater than 200 ms for the wild-type complex (Haqq et al., 1994).

Concluding Remarks

The stabilization of a distorted DNA conformation, as seen in the complex of hSRY–HMG with DNA, probably represents the underlying structural principle for the general function of HMG box proteins. For example, there have been reports demonstrating that HMG box proteins can functionally replace Escherichia coli host factors in in vitro recombination assays (Paull et al., 1993; Segall and Nash, 1994). This suggests that HMG domains can facilitate DNA bending in diverse environments. Further, members of the HMG-1/HMG-2 family, SRY in particular, have been shown to recognize fourway DNA junctions with high affinity and without any apparent sequence specificity (Ferrari et al, 1992). The structure of the complex of hSRY–HMG with duplex DNA suggests features of fourway DNA junctions such as a widened minor groove and underwound helix at the crossing point of the arms of the junction,

thereby providing a favorable binding site for an HMG domain.

The solution structure of the hSRY–HMG–DNA complex, in conjunction with the TBP–DNA cocrystal structure (Kim et al., 1993a, 1993b), demonstrates that sequence-specific recognition of a promoter-binding site represents only a portion of the activities necessary to regulate transcription. Although accomplished through very different three-dimensional (3D) folds and structural motifs, these two proteins, both of which bind exclusively in the minor groove, induce large conformational changes in the DNA, in particular helix unwinding, minor groove expansion, and DNA bending. This molecular switch can then bring distantly bound proteins of the transcription machinery into close proximity, thereby permitting them to interact with each other and influence transcription (Tjian and Maniatis, 1994). Interestingly, while TBP is found in all cell types, the sequence-specific HMG domains identified to date are associated with specific cell types (such as cells involved in sexual differentiation, in the case of SRY and SOX, and lymphocytes, in the case of LEF-1 and TCF-1). hSRY itself exerts a myriad of effects on male gonadal development. The structure of the hSRY–HMG–DNA complex presented in this paper permits the consequences of clinical mutations that result in 46X,Y sex reversal to be rationalized at the molecular level and suggests a common mechanism for the action of SRY on the expression of genes responsible for the male phenotype.

Experimental Procedures

Sample Preparation

The hSRY–HMG domain used in the present study comprises residues 57–133 of intact hSRY together with an N-terminal Met. In this paper, the N-terminal Met is numbered as residue 1. The hSRY–HMG domain, uniformly (>95%) labeled with either ^{15}N or ^{13}C and ^{15}N, was expressed in minimal medium using $^{15}NH_4Cl$, $^{13}C_6$–glucose, or both as the sole nitrogen and carbon sources and was purified as previously described (Ferrari et al., 1992) with minor modifications. The DNA strands were synthesized and purified as described (Omichinski et al., 1993). The complex was formed by mixing the purified components at room temperature in approximately equal proportions at a final protein concentration of 0.3 mg/ml. The dilute complex solution was then concentrated such that the samples used for the NMR experiments comprised 1.3 mM hSRY–HMG, 1.37 mM duplex DNA octamer, 10 mM sodium phosphate (pH 6.8), 5 mM NaN_3, and 50 mM EDTA in either 100% D_2O or 90% H_2O, 10% D_2O.

NMR Spectroscopy

All NMR experiments were carried out at 37°C on a Bruker AMX500 spectrometer equipped with a z-shielded gradient triple resonance probe. The sequential assignment of the ^1H, ^{13}C, and ^{15}N chemical shifts of hSRY–HMG in the complex was achieved by means of through-bond heteronuclear correlations along the backbone and side chains using the following 3D experiments: ^{15}N-separated HOHAHA, HNHA, CBCANH, CBCA(CO)NH, HBHA(CO)NH, C(CO)NH, HCCH–COSY, and HCCH–TOCSY. Details of these experiments, as well as the multidimensional NOE experiments, together with the original references, are provided in reviews by Bax and Grzesiek (1993) and by Clore and Gronenborn (1991, 1994). $^3J_{HN\alpha}$, $^3J_{CC}$, $^3J_{C\gamma N}$, $^3J_{C\gamma CO}$, $^3J_{NH\beta}$, and $^3J_{COH\beta}$ coupling constants were obtained by quantitative J correlation spectroscopy (Bax et al., 1994). Sequential assignments of the exchangeable and nonexchangeable protons of the DNA in the complex were obtained by standard procedures (Clore and Gronenborn, 1989) using 2D ^{12}C-filtered NOE (70 and 185 ms mixing times) and homonuclear Hartmann–Hahn spectra in D_2O and using a conventional 2D NOE spectrum (150 ms mixing time) in H_2O recorded with a 1-1 semi-selective excitation pulse. In addition, 2D ^{12}C-filtered rotating frame Overhauser enhancement (30 and 90 ms mixing times) spectra were recorded to assess the extent of spin–diffusion in the NOE spectra of the bound DNA. NOEs involving protons of the protein were obtained from 3D ^{15}N-separated, 3D ^{13}C-separated, and 4D $^{13}C/^{13}C$-separated NOE spectra (mixing times of 100, 120, and 150 ms, respectively). NOEs specifically between nonexchangeable protein and DNA protons were identified in a 3D ^{13}C-separated/^{12}C-filtered NOE spectrum (mixing time, 155 ms) recorded in D_2O.

Structure Calculations

Approximate interproton distance restraints were derived from the multidimensional NOE spectra, essentially as described previously (Omichinski et al., 1993). NOEs within the protein and between the protein and DNA 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. NOEs within the DNA were classified into five ranges: 1.8–2.5 Å, 1.8–3.0 Å, 1.8–3.5 Å, 2.3–5.0 Å, and 3.5–6.0 Å, corresponding to strong, medium-strong, medium, weak, and very weak NOEs, respectively. To account for the higher apparent intensity of methyl resonances, 0.5 Å was added to the upper distance limits for NOEs involving methyl protons. Distances involving methyl groups, aromatic ring protons, and nonstereospecifically assigned methylene protons were represented as a $(\Sigma r^{-6})^{-1/6}$ sum (Nilges, 1993). Protein backbone hydrogen-bonding restraints (two per hydrogen bond; $r_{NH\text{-}O}$ = 1.5–2.8 Å; $r_{N\text{-}O}$ = 2.4–3.5 Å) within areas of regular secondary structure were introduced during the final stages of refinement. Hydrogen-bonding restraints within the DNA were used to maintain base pairing (for GC base pairs, $r_{G(N1)\text{-}C(N3)}$ = 2.95 ± 0.2 Å; $r_{G(H1)\text{-}C(N3)}$ = 1.95 ± 0.2 Å; $r_{G(O6)\text{-}C(N4)}$ = 2.91 ± 0.2 Å; $r_{G(O6)\text{-}C(HN')}$ = 1.91 ± 0.2 Å; $r_{G(N2)\text{-}C(O2)}$ = 2.86 ± 0.2 Å; and $r_{G(HN')\text{-}C(O2)}$ = 1.86 ± 0.2 Å; for AT base pairs, $r_{A(N1)\text{-}T(N3)}$ = 2.82 ± 0.2 Å; $r_{A(N1)\text{-}T(H3)}$ = 1.82 ± 0.2 Å; $r_{A(N6)\text{-}T(O4)}$ = 2.95 ± 0.2 Å.; and $r_{A(HN')\text{-}T(O4)}$ = 1.95 ± 0.2 Å; Saenger, 1984). Torsion angle restraints of 71 ϕ, 10 ψ, 56 χ^1, and 16 χ^2 angles were derived from the NOE and coupling constant data, and the minimum ranges employed were ±10°, ±50°, ±20°, and ±30°, respectively (Clore and Gronenborn, 1994). We also employed 72 broad torsion angle restraints, covering the values characteristic for both A- and B-DNA, for the DNA backbone in order to prevent problems associated with local mirror images (Omichinski et al., 1993): α = 60° ± 50°, β = 180° ± 50°, γ = 60° ± 35°, ε = 180° ± 50°, and ζ = −85° ± 50°. The structures were calculated using a modified version of the hybrid distance geometry dynamical–simulated annealing protocol (Nilges et al., 1988) using the program XPLOR-31 (Brünger, 1993), adapted to incorporate pseudopotentials for $^3J_{HN\alpha}$ coupling constant (Garrett et al., 1994) and secondary $^{13}C\alpha$ and $^{13}C\beta$ chemical shift (Kuszewski et al., 1995) restraints. The target function that is minimized during simulated annealing and restrained regularization comprises only quadratic harmonic potential terms for covalent geometry, $^3J_{HN\alpha}$ coupling constant and secondary $^{13}C\alpha$ and $^{13}C\beta$ chemical shift restraints, square-well quadratic potentials for the experimental distance and torsion angle restraints, and a quartic van der Waals repulsion term for the nonbonded contacts. There were no hydrogen-bonding, electrostatic, or 6-12 Lennard–Jones empirical potential energy terms in the target function.

All the structural DNA parameters were calculated with the program CURVES (Lavery and Sklenar, 1989). The width of the minor groove is defined as the minimum O4′–O4′ distance across the groove minus 2.8 Å (for the sum of the van der Waals radii of two oxygen atoms). Structures were displayed with the programs AVSXPLOR (Brünger and DeLano, 1993), VISP (de Castro and Edelstein, 1992), and GRASP (Nicholls, 1993).

The coordinates of the 35 final simulated annealing structures of the specific complex of the hSRY–HMG with DNA, together with the coordinates of the restrained regularized mean structure, $\langle\overline{SA}\rangle r$, and the complete list of experimental NMR restraints and ^1H, ^{15}N, and ^{13}C assignments have been deposited in the Brookhaven Protein Data Bank.

Acknowledgments

Correspondence should be addressed to A. M. G. and G. M. C. We

thank M. Bianchi for providing us with the hSRY–HMG clone and for useful discussions; L. Trinh and J. Shiloach for help with the bacterial expression; D. Garrett and F. Delaglio for software support; R. Tschudin for technical support; E. de Castro and S. Edelstein for the program VISP; and A. Bax, R. Clubb, A. Wang, and S. Grzesiek for numerous stimulating 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. and A. M. G.).

Received March 10, 1995; revised March 30, 1995.

References

Affara, N. A., Chalmers, I. J., and Ferguson-Smith, M. A. (1993). Analysis of the *SRY* gene in 22 sex-reversed XY females identifies four new point mutations in the conserved DNA binding domain. Hum. Mol. Genet. *2*, 785–789.

Bax, A., and Grzesiek, S. (1993). Methodological advances in protein NMR. Acct. Chem. Res. *26*, 131–138.

Bax, A., Vuister, G. W., Grzesiek, S., Delaglio, F., Wang, A. C., Tschudin, R., and Zhu, G. (1994). Measurement of homo- and heteronuclear J couplings from quantitative J correlation. Meth. Enzymol. *239*, 79–106.

Berta, P., Hawkins, J. R., Sinclair, A. H., Taylor, A., Griffiths, B., Goodfellow, P. N., and Fellous, M. (1990). Genetic evidence equating *SRY* and the testis-determining factor. Nature *348*, 448–450.

Braun, A., Kammerer, S., Cleve, H., Löhrs, U., Schwarz, H.-P., and Kuhnle, U. (1993). True hermaphroditism in a 46X,Y individual, caused by a postzygotic somatic point mutation in the male gonadal sex-determining locus (*SRY*): molecular genetics and histological findings in a sporadic case. Am. J. Hum. Genet. *52*, 578–585.

Brooks, B. R., Bruccoleri, R. E. Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983). CHARMM: a program for macromolecular energy minimization and dynamics calculations. J. Comput. Chem. *4*, 187–217.

Brünger, A. T. (1993). XPLOR Manual, Version 3.1 (New Haven, Connecticut: Yale University).

Brünger, A. T., and DeLano, W. (1993). AVS-XPLOR User's Manual (New Haven, Connecticut: Yale University).

Clore, G. M., and 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.

Clore, G. M., and Gronenborn, A. M. (1991). Structures of larger proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy. Science *252*, 1390–1399.

Clore, G. M., and Gronenborn, A. M. (1994). Structures of larger proteins, protein–ligand and protein–DNA complexes by multidimensional hetronuclear NMR. Prot. Sci. *3*, 372–390.

de Castro, E., and Edelstein, S. (1992). VISP 1.0 User's Guide (Geneva, Switzerland: University of Geneva).

Eisenberg, D., and McLaghlan, A. D. (1986). Solvation energy in protein folding and binding. Nature *319*, 199–203.

Ferrari, S., Harley, V. R., Pontiggia, A., Goodfellow, P. N., Lovell-Badge, R., and Bianchi, M. E. (1992). SRY, like HMG1, recognizes sharp angles in DNA. EMBO J. *11*, 4497–4506.

Garrett, D. S., Kuszewski, J., Hancock, T. J., Lodi, P. J., Vuister, G. W., Gronenborn, A. M., and Clore, G. M. (1994). The impact of direct refinement against three-bond HN–CαH coupling constants on protein structure determination by NMR. J. Magn. Reson. (B) *104*, 99–103.

Giese, K., Cox, J., and Grosschedl, R. (1992). The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. Cell *69*, 185–195.

Goodfellow, P. N., and Lovell-Badge, R. (1993). SRY and sex determination in mammals. Annu. Rev. Genet. *27*, 71–92.

Gustafson, M. L., and Donahoe, P. K. (1994). Male sex determination: current concepts of male sexual differentiation. Annu. Rev. Med. *45*, 505–524.

Haqq, C. M., King, C.-Y., Donahoe, P. K., and Weiss, M. A. (1993). 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., and Weiss, M. A. (1994). Molecular basis of mammalian sexual determination: activation of Müllerian inhibiting substance gene expression by SRY. Science *266*, 1494–1500.

Harley, V. R., Jackson, D. I., Hextall, P. J., Hawkins, J. R., Berkovitz, G. D., Sockanathan, S., Lovell-Badge, R., and Goodfellow, P. N. (1992). DNA binding activity of recombinant SRY from normal males and XY females. Science *255*, 453–456.

Harley, V. R., Lovell-Badge, R., and Goodfellow, P. N. (1994). Definition of a consensus DNA binding site for SRY. Nucl. Acids Res. *22*, 1500–1501.

Hawkins, J. R., Taylor, A., Goodfellow, P. N., Migeon, C. J., Smith, K. D., and Berkovitz, G. D. (1992a). Evidence for increased prevalence of *SRY* mutations in XY females with complete rather than partial gonadal dysgenesis. Am. J. Hum. Genet. *51*, 979–984.

Hawkins, J. R., Taylor, A., Berta, P., Levilliers, J., Van der Auwera, B., and Goodfellow, P. N. (1992b). Mutational analysis of *SRY*: nonsense and missense mutations in XY sex reversal. Hum. Genet. *88*, 471–474.

Jones, D. N. M., Searles, M. A., Shaw, G. L. Churchill, M. E. A., Ner, S. S., Keeler, J., Travers, A. A., and Neuhaus, D. (1994). The solution structure and dynamics of the DNA-binding domain of HMG-D from *Drosophila melanogaster*. Structure *2*, 609–627.

Kim, Y., Geiger, J. H., Hahn, S, and Sigler, P. B. (1993a). Crystal structure of a yeast TBP/TATA-box complex. Nature *365*, 512–520.

Kim, J. L., Nikolov, D. B., and Burley, S. K. (1993b). Co-crystal structure of TBP recognizing the minor groove of a TATA element. Nature *365*, 520–527.

King, C.-Y., and Weiss, M. A. (1993). The SRY high-mobility-group box recognizes DNA by partial intercalation in the minor groove: a topological mechanism of sequence specificity. Proc. Natl. Acad. Sci. USA *90*, 11990–11994.

Kuszewski, J., Qin, J., Gronenborn, A. M., and Clore, G. M. (1995). The impact of direct refinement against $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts on protein structure determination by NMR. J. Magn. Reson. (B) *106*, 92–96.

Laudet, V., Stehelin, D., and Clevers, H. (1993). Ancestry and diversity of the HMG box superfamily. Nucl. Acids Res. *21*, 2493–2501.

Lavery, R., and Sklenar, H. (1989). Defining the structure of irregular nucleic acids: conventions and principles. J. Biomolec. Struct. Dyn. *6*, 655–667.

McElreavy, K. K., Vilain, E., Abbas, N., Costa, J.-M., Souleyreau, N., Kucheria, K., Boucekkine, C., Thibaud, E., Brauner, R., Flamant, F., and Fellous, M. (1992). XY sex reversal associated with a deletion 5′ to the SRY "HMG box" in the testis-determining region. Proc. Natl. Acad. Sci. USA *89*, 11016–11020.

Nicholls, A. J. (1993). GRASP Manual (New York: Columbia University).

Nilges, M. (1993). A calculational strategy for the structure determination of symmetric dimers by ^{1}H NMR. Prot. Struct. Funct. Genet. *17*, 295–309.

Nilges, M., Clore, G. M., and Gronenborn, A. M. (1988). Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing. FEBS Lett. *229*, 129–136.

Omichinski, J. G., Clore, G. M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S. J., and Gronenborn, A. M. (1993). NMR structure of a specific DNA complex of a Zn-containing DNA binding domain of GATA-1. Science *261*, 438–446.

Paull, T. T., Haykinson, M. J., and Johnson, R. C. (1993). The nonspecific DNA-binding and bending proteins HMG1 and HMG2 promote the assembly of complex nucleoprotein structures. Genes Dev. *7*, 1521–1534.

Pontiggia, A., Rimini, R., Harley, V. R., Goodfellow, P. N., Lovell-Badge, R., and Bianchi, M. E. (1994). Sex-reversing mutations affect the architecture of SRY–DNA complexes. EMBO J. *13*, 6115–6124.

Read, C. M., Cary, P. D., Crane-Robinson, C., Driscoll, P. C., and Norman, D. G. (1993). Solution structure of a DNA-binding domain from HMG-1. Nucl. Acids Res. *21*, 3427–3436.

Read, C. M., Cary, P. D., Preston, N. S., Lnenicek-Allen, M., and Crane-Robinson, C. (1994). The DNA sequence specificity of HMG boxes lies in the minor wing of the structure. EMBO J. *13*, 5639–5646.

Saenger, W. (1984). Principles of Nucleic Acid Structure (New York: Springer-Verlag).

Segall, A. M., and Nash, H. A. (1994). Architectural elements in nucleoprotein complexes: interchangeability of specific and non-specific DNA binding proteins. EMBO J. *13*, 4536–4548.

Sinclair, A. H., Berta, P., Palmer, M. S., Hawkins, J. R., Griffiths, B. L., Smith, M. J., Foster, J. M., Frischauf, A.-M., Lovell-Badge, R., and Goodfellow, P. N. (1990). A gene from the human sex determining region encodes a protein with homology to a conserved DNA binding motif. Nature *346*, 240–244.

Tajima, T., Nakae, J., Shinohara, N., and Fujieda, K. (1994). A novel mutation localized in the 3′ non-HMG box region of the *SRY* gene in 46X,Y gonadal dysgenesis. Hum. Mol. Genet. *3*, 1187–1189.

Tjian, R., and Maniatis, T. (1994). Transcriptional Activation: a complex puzzle with few easy pieces. Cell *77*, 5–8.

Tucker, P. K., and Lundrigan, B. L. (1993). Rapid evolution of the sex determining locus in Old World mice and rats. Nature *364*, 715–717.

van de Wetering, M., and Clevers, H. (1992). Sequence-specific interaction of the HMG box proteins TCF-1 and SRY occurs within the minor groove of a Watson–Crick double helix. EMBO J. *11*, 3039–3044.

Weir, H. M., Kraulis, P. J., Hill, C. S., Raine, A. R. C., Laue, E. D., and Thomas, J. O. (1993). Structure of the HMG box motif in the B-domain of HMG-1. EMBO J. *12*, 1311–1319.

Whitfield, L. S., Lovell-Badge, R., and Goodfellow, P. N. (1993). Rapid sequence evolution of the mammalian sex-determining gene *SRY*. Nature *364*, 713–715.