Published in final edited form as:

J Mol Biol. 2009 January 16; 385(2): 469–490. doi:10.1016/j.jmb.2008.10.041.

Crystallographic Analysis of a Sex-Specific Enhancer Element: Sequence-Dependent DNA Structure, Hydration, and Dynamics

Narendra Narayana^{*,†} and Michael A Weiss

Department of Biochemistry, Case Western Reserve University, Cleveland, Ohio 44106

Abstract

The crystal structure of a sex-specific enhancer element is described at a resolution of 1.6 Å. This 16-base-pair site, designated *Dsx^A*, functions in the regulation of a genetic switch between male- and female patterns of gene expression in *Drosophila melanogaster*. Related sites are broadly conserved in metazoans, including in the human genome. This enhancer element is unusually rich in general regulatory sequences related to DNA recognition by multiple classes of eukaryotic transcription factors, including the DM motifs, homeodomain, and high-mobility-group (HMG) box. Whereas free DNA is often crystallized as an A-form double helix, *Dsx^A* was crystallized as B-DNA and so provides a model for the pre-bound conformation of diverse regulatory DNA complexes. Sequence-dependent conformational features are observed that extend features of shorter B-DNA fragments with respect to double-helical parameters, groove widths, hydration, and binding of divalent metal ions. The structure also exhibits a sequence-dependent pattern of isotropic thermal *B*-factors, suggesting possible variation in the local flexibility of the DNA backbone. Such fluctuations are in accord with structural variability observed in prior B-DNA structures. We speculate that sites of intrinsic flexibility within a DNA control element provide hinges for its protein-directed reorganization in a transcriptional pre-initiation complex.

Keywords

DM domain; DNA structure; Doublesex; protein-DNA recognition; sex determination

Introduction

DNA structure is fundamental to diverse DNA transactions, such as replication, repair, nucleosome assembly, and the regulation of gene expression.^{1,2} Architectural and regulatory proteins bind to DNA with an affinity, specificity, and kinetics appropriate to their biological roles.³ Comparative structural analyses of DNA oligonucleotides and protein-DNA complexes have provided insight into the physical principles of DNA structure, reorganization, and recognition. DNA-protein recognition is enabled by the characteristic structure of the double helix, including base-specific arrays of functional groups in the major- and minor grooves,⁴ specific water molecules,⁵ and counter-ion distributions.^{6,7} Of particular interest is the potential contribution of DNA dynamics and induced fit^{8–11} to recognition: should specific B-

*Corresponding author. E-mail: narayanan3@uthscsa.edu.

†Present address: Department of Biochemistry, The University of Texas, Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, Texas 78229-3900, USA

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

DNA sequences differ in malleability,^{12,13} such rigid or flexible structural elements may contribute to “indirect read-out” of the underlying DNA sequence.^{5,14,15} In this Article we describe at high resolution the crystal structure of an enhancer DNA element that functions in a well-characterized genetic switch, the reciprocal regulation of male- and female gene expression by the Doublesex proteins in *Drosophila melanogaster*.¹⁶ Designated *dsx^A*, this enhancer element contains subsites related to multiple classes of eukaryotic transcription factors (Table 1). Because in most cases structures of protein-DNA complexes have been determined in the absence of the structures of their cognate free DNA sites, the structure of *dsx^A* provides a model for general analysis of sequence-specific recognition and induced fit by diverse DNA-binding motifs. Our results extend previous crystallographic studies of sequence-dependent structural features in shorter oligonucleotides and provide novel evidence for the sequence-dependent flexibility of B-DNA at discrete sites of local backbone disorder.

Because of their fundamental importance, crystallographic studies of DNA fragments have been pursued for over two decades. Much of our understanding of sequence-dependent features of free B-DNA structures stems from extensive structural studies of the Dickerson-Drew dodecamer (DDD; d(CGCGAATTCGCG)₂) and related sequence isomers and decamers.¹² Throughout the text, references to *DDD* or *R3 dodecamer crystal form* relate to the sequence d(CGCGAATTCGCG)₂. These structures exhibit sequence-dependent conformations with respect to base-pair geometry, backbone flexibility, groove widths and depths, hydration, metal ion coordination, bending and deformability.^{12,17} Two key features are variation of the width of the minor groove and bending of the DNA double helix. Such conformational variations contribute to sequence-specific recognition of DNA by proteins. In addition to geometric features of the double helix, the binding of water molecules and cations at specific sites is influenced by the functional groups of DNA bases.⁶ Although the positions and types of cations surrounding the DNA duplex can be important determinants of lattice packing and influence local DNA conformation,⁷ the broad relevance of crystallographic features to the structure of DNA in solution is well supported by NMR studies based on nuclear Overhauser effects¹⁸ and residual dipolar couplings.¹⁹

We describe herein the crystal structure of an A+T-rich DNA duplex (5'-ACACTACAATGTTGCAAT-3' and complement; top panel Figure 1) derived from the *Drosophila fat body enhancer*.²⁰ The central segment (underlined) defines *dsx^A*, a control site that contributes to the sex- and tissue-specific expression of yolk proteins by the male- and female isoforms of the Doublesex transcription factor.²⁰ This structure (16 base pairs; Figure 1, bottom panel) represents the longest free B-DNA duplex structure to be characterized at atomic resolution (1.6 Å) to date, sufficient to visualize sequence-dependent variations in structure, hydration, and metal-ion binding. The sequence is notable for a central A+T-rich segment (5'-AATGTT-3') critical to recognition by the conserved DM family of minor-groove DNA-binding proteins.^{21,22} Although the DNA site is not self-complementary, the structure exhibits a striking pseudo-2-fold symmetry with respect to base-pair and sugar-phosphate geometries, groove widths, hydration patterns, and metal-cation distribution. These parameters appear to be context- and sequence- specific. The central portion of the *dsx^A* site (5'-AATGTT-3') is straight with B-type helical parameters consistent with previous A+T-rich B-DNA structures;²³ the adjoining three base-pair steps on either side resemble an alternating B-type conformation.²⁴ The terminal two base-pair steps at each end are slightly bent relative to the central site. Because subsites within *dsx^A* resemble target sites for diverse classes of eukaryotic transcription factors contacting either or both DNA grooves (Table 1), their structural features within free DNA may be of broad biological interest.

An unexpected feature of *dsx^A* relates to its pattern of crystallographic order and disorder. Whereas variations in mean atomic coordinate position are consistent with the database of DNA structures,²³ thermal *B*-factors exhibit anomalies. Specific deoxyribose moieties within

the central A+T-rich segment (5'-AAT**G**TT-3' (top) and complement 3'-TTACAA-5'; bold) exhibit unusually weak electron density (Figures 2B and 2C), suggesting conformational fluctuations within an otherwise well-defined DNA segment (Figure 2A). The electron density of a specific phosphodiester bridge (in bottom strand 3'-TTA[↓]CAA-5'; Figure 2B) is likewise ill-defined. These anomalies appear unrelated to effects of lattice packing and occur at sites distant from the ends of the oligonucleotides. We therefore propose that the disordered sites within the *dsx*^A enhancer element are intrinsic to its sequence. Sequence-dependent variations in DNA malleability may contribute to its reorganization in DNA-protein assemblies.

Results

Gel mobility-shift assays demonstrate that the DM domain of Doublesex and the unrelated HMG box of human SRY can each bind *dsx*^A with high affinity. In the crystal the DNA duplex forms 1.5 turns of a B-DNA. The central subsite 5'-AAT**G**TT-3' (and complement) resembles in sequence a canonical AT-tract 5'-AATTTT-3' with one T→G transversion (bold). This analogy corresponds to structural similarities as described in detail below. We therefore designate the central 5'-AAT**G**TT-3' sub-site as pseudo-AT-tract. Like a canonical AT-tract, this sub-site is straight with high propeller twist, narrow minor groove, and hydration spine. Although the G-C substitution can thus be accommodated within this DNA motif, it is associated with higher temperature factors (*B*-factors) and anomalies at the T10-G11/C28-A29 step.

DNA Structure

In the 16-base-pair DNA duplex considered as whole, the average helical twist between successive base-pairs is 33.7°, corresponding to 10.7 residues per turn. This value is close to that observed in solution for random DNA and alternating poly(dA)·poly(dT),^{25,26} but differs from the mean helical-repeat length observed in the majority of DNA dodecamer crystals (generally between 10.0 and 10.2 base pairs per turn). The present value is nonetheless similar to that recently seen in R3 crystal form of DDD (free) and the crystal structure of a specific AT-rich DNA duplex (5'-AGGTGTGAAATT-3' and complement) bound to the T domain of the *Brachyury* transcription factor.²⁷ Unlike the latter complex, which exhibits a widened minor groove, in the present structure the minor groove is narrow and deep. Conversely, the major groove is wide and shallow. The inner 12-base pairs (5'-TACAATGTTGCA-3' and complement) are nearly straight with a slight bend toward the major groove at the flanking residues. The 16-base-pair DNA site exhibits a mean pitch of 3.3 Å and an average diameter of 18.8 Å; these values are within the range expected of B-DNA. The average thermal *B*-factors for bases, sugars, and phosphates (each considered as a group) are 13, 16, and 19 Å², respectively (Figure 3); a similar trend has been observed in other DNA structures. Such non-uniform distribution of *B*-factors across the helical radius may reflect a combination of static disorder, librational disorder, and local thermal motions.^{28,29} The double helices stack on one another along the crystallographic *c*-axis, mimicking a continuous DNA duplex in the lattice (Figure 4B).

A striking deviation from a classical model of B-DNA (as derived from fiber diffraction data³⁰) arises from variation in the minor-groove width. Whereas in the fiber model the width of B-DNA minor groove is 5.7 Å, the present structure exhibits significant variation in minor-groove widths; the pattern is quasi-symmetrical about the T12-T13 step. Beginning from steps T12-T13 on one strand and T30-T31 on the other strand, the minor groove opens gradually towards both the ends of the double helix, varying from 4.1 Å (phosphorus atoms P12-P31 and P13-P30) to 7.3 Å (P7-P36 and P18-P25). The value for groove width corresponds to P---P distance between the appropriate phosphate groups on the two strands minus 5.8 Å for the radii of phosphate groups. The mean minor-groove width is 5.8 Å (± 1.2 Å), similar to that of the

fiber-diffraction model. This marked narrowness is a general feature of a canonical A-tract³¹ and A_mT_n-tract B-DNA structures (m, n ≥ 2, typically with m = n). The narrowed minor groove was previously found in the canonical and related dodecamers, as well as in A-tract containing DNA sequences.³²⁻³⁴ The characteristic sequence-dependent minor groove narrowness in the A/T region is markedly symmetric as in the trigonal and rhombohedral forms of d (CGCGAATTCGCG)₂.^{35,7} The mean width and depth values (Å) for the minor and major grooves,³⁶ are 5.5 (1.0), 5.4 (0.6), and 12.4 (0.7), 2.4 (2.3), respectively. Values within parentheses denote standard deviations.

Sugar-phosphate backbone angles, glycosyl torsion angles, and pseudo-rotation parameters exhibit expected trends (Supporting Information) with specific departures as noted below. The majority of the deoxyribose rings (20 of 32 observed in the structure) adopt standard C2'-*endo* geometry whereas a few are C1'-*exo*. Torsion angles α, β, γ, and δ are in their usual (-)-*gauche*, *trans*, (+)-*gauche*, and (+)-*anticlinal* ranges, respectively. Exceptions are found for the β angles of G14, A16, G32, and A34, which assume (+)-*anticlinal* conformation. 24 of 30 phosphodiester moieties exhibit B_I conformations. For this set the mean ε and ζ dihedral angles are -179° ± 7° and -99° ± 9°, respectively. Glycosidic angles (χ) are all (-)-*anticlinal*, with an average of -102° ± 12°. Six of 30 phosphodiester groups (C7, T13, and C15 on strand 1, and C25, T31, and T33 on strand 2) adopt B_{II} conformations. Although observation of a small subset of such B_{II} conformations is not uncommon,³⁷ the present structure is unusual in that the four of the six B_{II} sites (T13, C15, T31, and T33) are in or near the narrow minor groove (widths 4.1-5.4 Å) of the central 5'-ACAATGTT-3' segment; B_{II} phosphodiester conformations are ordinarily associated with widening of the minor groove.³⁸ The narrow minor groove in the central AATT region in the DDD is free from B_{II} conformation. Indeed, the remaining two B_{II} phosphodiester groups (at C7 and C25) adjoin such a widened minor groove (width ~ 7.3 Å). Although the *dsx*^A DNA duplex is non-palindromic, corresponding torsion angles in the backbones of strands 1 and 2 exhibit a pseudo two-fold symmetry.

The DNA structure exhibits marked sequence-specific variation in local base-pair geometry. Local helical parameters, as determined by the method of Olson and coworkers³⁹, are given in Table 2. In each case bases were unambiguously defined in the Fourier maps (Figure 2), enabling precise characterization of these parameters. Helical twist is uniform in the 5'-AATGTT-3' central segment and varies in a low-high-low pattern in the flanking regions. Twist values exhibit a near two-fold symmetrical pattern about the central T10-G11 step. Steps A6-C7 and G14-C15 exhibit the lowest twist angles (21.8° and 21.5°, respectively) whereas the highest twist angles occur at base-pair steps C7-A8 (47.9°) and T13-G14 (49.1°). The pattern of roll angles is remarkable for a regular low-high alternation along the helical axis, likewise with approximate two-fold symmetry about T10-G11 base pair step. Roll angles in the central 5'-CAATGTTG-3' segment are small with a maximum value of 8.5° at the T10-G11 step and a minimum value of -6.0° at the T13-G14 step. The overall mean value for roll angles is 2.0° (± 3.7°). Interestingly, roll angles in the A6-C15 region alternate between positive and negative values. The terminal three base-pair steps at each end of the DNA duplex exhibit small successive positive roll angles (~3°), resulting in local compression of the major groove and minor DNA bending. As a consequence, whereas the DNA bend angle of the central 12 base-pair segment is only 4°, the overall bend angle of the 16 base-pair duplex is 15°. ⁴⁰ The bends occur at the T5-A6 step on one end and at T23-G24 step at the other. Considering the general observation that pyrimidine-purine steps are susceptible for bending, particularly C-A/T-G steps, it is noteworthy that C7-A8/T31-G32 and T10-G11/C28-A29 base-pair steps within the A+T-rich *dsx*^A site do not exhibit distortions. As observed among other geometric parameters, the pitch of the DNA double helix displays a near two-fold symmetrical distribution; values are clustered in a narrow range (3.0-3.6 Å). The mean axial rise per base pair is 3.3 Å (± 0.1 Å), which is less than that of canonical B-DNA models based on fiber diffraction (3.4 Å). The T10-G11 step interrupting the central A_mT_n-tract exhibits a maximum rise of 3.6 Å.

Base pairs within the core of *dsx*^A site exhibit the expected range of non-coplanarity due to propeller twist and buckle. Propeller twists for each of the 16 base pairs are negative with a near two-fold symmetry about an orthogonal axis passing between T10 and G11. The mean propeller twist value is $-12^\circ (\pm 5^\circ)$. Larger values of propeller twist occur within the six central base pairs (5'-AATGTT-3' and complement; average twist -15°); the lone G-C base pair in this segment (bold) exhibits a high propeller twist similar to those of the neighboring A-T base pairs. By contrast, all other C-G base pairs in the structure exhibit little propeller twist with the exception of C4-G35 at the 5'-end (-14°). A-T base pairs assume either low or high propeller twists in the flanking region. Base pairs A6-T33 and C7-G32 (preceding the central 5'-AATGTT-3' segment) do not exhibit high propeller twists. Buckle values vary widely from -15° to $+13^\circ$ with a mean value of $-1^\circ (\pm 9^\circ)$. Uniformly strong buckling occurs in the central 5'-AATGTT-3' segment with positive values in sub-segment A8-A9-T10/A29-T30-T31 and negative values in sub-segment G11-T12-T13/A26-A27-C28. Such variations correspond to a near inversion symmetry about T10-G11 base pair step, distinct from the near-two-fold symmetry displayed by other parameters. Base-pair opening in *dsx*^A is consistently positive and generally larger in the central 5'-AATGTT-3' segment relative to the flanking segments. The mean opening angle is $2^\circ (\pm 2^\circ)$ with a maximum at base pair T10-A29 (4.2°), which adjoins the “interrupting” G-C base pair in the central segment. Variation in slide (restricted to the range -0.8 and 2.2 Å; Table 2) in *dsx*^A is negligible; the mean slide distance is 0.1 Å (± 1.0 Å).

Hydration

197 water molecules have been identified in the asymmetric unit, of which 82 are in the minor groove and 115 in the major groove. In accord with other structures, the narrow minor groove of the central AT-rich segment exhibits two well-organized layers of water molecules to form a zig-zag spine (Figures 5 and 6). This spine is comprised of a primary shell (seven water molecules) and a secondary shell (four water molecules). The water molecules in the first shell are hydrogen-bonded to pyrimidine O2 atoms and purine N2 and N3 atoms. Single water molecules mediate cross-strand hydrogen-bonding interactions between bases in adjacent base pairs, with the exception of the ends of the spine where two water molecules link N3 and O2 atoms (A9→T31 and T13→A27; Figure 5). The presence of a C-G base pair in the central 5'-AATGTT-3' segment (bold) does not perturb the hydration spine. On either side of the central segment, hydrated calcium ions are found in the widened minor groove (see below), and two linear arrays of water molecules face the DNA strands (Figure 5). Thus, sequence-specific broadening of the minor groove and binding of calcium ions, at each end of the DNA duplex limits extension of the central spine of hydration. Cross-strand bifurcated hydrogen bonds in the major groove (of the type described in DNA A-tracts³⁴) are not prominent. Such bifurcated interactions can only occur at sequences of two or more successive adenines. In our case the distances A8:N6 - 30T:O4 (3.4 Å) and 26A:N6 - 12T:O4 (3.5 Å) on the major groove of the central A-T-rich core are long, suggesting weak or absence of hydrogen bond.

Calcium-binding Sites

The DNA site was crystallized in the presence of calcium ions. The electron-density map at 1.6 Å resolution enabled identification of 12 independent Ca^{2+} ions per asymmetric unit (Supporting Information). There are thus 20 Ca^{2+} ions (including symmetry-related partners) in the immediate surroundings of the DNA duplex (Figure 4A). This represents the largest number of Ca^{2+} ions bordering a DNA duplex whose structure has been determined by X-ray crystallography. Of the 12 unique Ca^{2+} ions, four reside in the minor groove with unit occupancy (designated Ca37-Ca40; Figures 4A and 6B), and the remaining eight Ca^{2+} ions (Ca41-Ca44 in one subset and Ca45-48 in another) are located on crystallographic 3-fold symmetry axes at the periphery of the duplex but close to the non-phosphodiester oxygens. As noted above in regard to several geometric parameters, the substructure of calcium ions displays

a near two-fold symmetry about an orthogonal axis passing between base pairs T10-A29 and G11-C28. Calcium ions in the minor groove are octa-coordinated (Figures 7A and 7B) with ligand distances ranging between 2.2 and 2.9 Å. Those that lie on three-fold symmetry axes are hexa-coordinated (Figures 7C and 7D) with coordination distances close to 2.3 Å. The calcium ions in the minor groove interact with the base atoms through inner-shell water molecules. Of these, two ions (Ca39 and Ca40) are present on either side of the central spine of hydration but within the interior of the minor groove; these adjoin base-pair steps T13-G14 and C7-A8, respectively. Ca37 and Ca38 are by contrast at the ends of the minor groove at base-pair steps A16-A17 and C4-T5, respectively.

The calcium ions that lie on a three-fold symmetry axes are classified into two subsets, Ca41-Ca44 and Ca45-Ca48. Ca41-Ca44 interact with three DNA duplexes related by a symmetry axis projecting laterally through the non-bridging oxygen atoms O1P and O2P of residues A8 and A26. Inner-sphere coordination of these calcium ions is mediated in part by the above non-bridging oxygens and in part by specific bound water molecules (Figure 7C). The second set of calcium ions (Ca45-Ca48) pertains to the four termini of the DNA duplex. One is observed at each terminus near a symmetry-related partner. These pairs connect one group of three DNA duplexes in the crystal lattice (see below) with neighboring sets of three DNA duplexes along the helical axis, thus forming calcium-bridged continuous pillars of helices (Figure 7D). To our knowledge, this is the first example of such calcium bridging. The calcium ions in the second set are likewise surrounded by six coordinating oxygen atoms, either belonging to water molecules or non-bridging oxygens of the DNA backbone.

Crystal packing

In the crystal lattice the DNA sites form infinite columns along the crystallographic *c*-axis with continuous major- and minor grooves. The overall packing scheme contains three duplexes related by $\pm 120^\circ$ rotations, stacked successively on top of one another along the *c*-axis. The DNA backbone is continuous with the exception of the missing 5'-terminal phosphate groups (Figure 4B). Terminal base pair T18-A21 of one DNA site is thus stacked on terminal base pair A3-T36 of another DNA site related by a screw symmetry axis; the rise between these base pairs is 3.8 Å, longer than the mean rise distance of 3.3 Å within the DNA site. Stacking of these terminal base pairs is characterized by a negative inter-helical twist (*ca.* -25.5°). At this interface the minor groove is enlarged and contains several water molecules. Helical stabilization by metal ions along the *c*-axis is provided by the shared Ca^{2+} ion pairs located at the ends of the strands as described above. The lateral calcium ions appear to shield electrostatic repulsion between side-by-side DNA sites (Figures 4B and 7C). These divalent cations bind at the confluence of partial negative charges interposed by non-bridging oxygen atoms of the DNA backbone.

Sequence-Dependent DNA Dynamics and Disorder

Overall relative isotropic thermal factors ($\text{\AA}^2$) are: all DNA atoms, 15; calcium ions, 16; and water molecules, 27. Mean *B*-values ($\text{\AA}^2$) for phosphate groups, deoxyribose sugars, and bases are 19, 16, and 13, respectively (Figure 3). In general, the 5'-AATGTT-3' segment exhibits larger *B*-values than the flanking sequences. Of specific interest is that the central nucleotides exhibit higher *B*-factors relative to the mean value of 15 $\text{\AA}^2$: T10 (23 $\text{\AA}^2$) and G11 (21 $\text{\AA}^2$) and the corresponding complement A29 (21 $\text{\AA}^2$) and C28 (24 $\text{\AA}^2$). In $F_o - F_c$ and $2F_o - F_c$ maps, the deoxyribose rings of residues T10, C28, and G11 display relatively poor density ($< 0.5 \sigma$; Figures 2B and 2C). Although it is generally difficult to distinguish effects of crystal packing from intrinsic properties of a macromolecule, we propose that the above variations in *B*-factors and the associated poor density are sequence-dependent features. This perspective is supported by three lines of reasoning. First, the present DNA duplex is sufficiently long that the central *dsx*^A sub-site is well insulated from end- or packing effects (i.e., by four base pairs on either

side). Such insulation is in accord with classical observations by Dickerson and colleagues that the geometrical parameters associated with a specific DNA sequence flanked on either side by four base pairs may be considered to be intrinsic.⁴¹ Our analysis suggests that this notion may be extended to thermal parameters. Second, in the present lattice individual DNA duplexes exhibit only sheer stacking interactions along the *c*-axis. Such peripheral stacking between neighboring duplexes represents a less extensive lattice contact than that observed in DDD structures, which typically exhibit the rigorous interlocking of neighboring duplexes in the minor groove via the two terminal residues. The canonical four-base-pair insulation, although present in the present structure, would thus appear by contrast to DDD to be less needed as a barrier from lattice contacts. Third, the anomalous nucleoside moieties of the central *dsx*^A sub-site are not involved in inter-duplex interactions. A plausible model for conformational fluctuations underlying the anomalous *B*-factors in the central sub-site can be visualized by molecular dynamics (see Discussion and Supporting Information). Alternatively, the relatively lower *B*-factors for the terminal residues compared to the central A-T-rich region may be due to effects of crystal packing interactions, including calcium coordination.

Discussion

The present studies provide crystallographic evidence for the inherent flexibility of specific base pair steps within a B-DNA duplex. Such malleable sites in a chromosome may function as hinges in protein-directed reorganization of DNA, a general feature of diverse DNA transactions in a cell. We discuss the structure and dynamics of *dsx*^A with possible implications for binding of homeodomains and the DM domain of Doublesex.

Relationship to Homeodomain-DNA Recognition

The sequence of *dsx*^A contains subsites related to homeodomain-DNA recognition (Table 1). Recent studies by Joshi *et al.*,⁴² demonstrate that sequence-dependent DNA structure may play a pivotal role in conferring specificity to a Hox protein at a key step in developmental patterning. In *D. melanogaster* a specific homeotic gene product (the Sex combs reduced (Scr) protein) binds cooperatively with its homeodomain-containing cofactor Extradenticle (Exd) to a paralog-specific DNA site (5'-AGATTAATCG-3' and complement). Based on comparison of two ternary structures, wherein the same protein heterodimer (Scr-Exd) is bound either to the above site or a Hox consensus site (5'-TGATTTATGG-3' and complement), Joshi *et al.*, observed that specific minor-groove insertion of an Arg-His recognition element in the N-terminal arm of Scr occurs only in the paralog-specific site. In contrast the two complexes share similar contacts between the Scr homeodomain recognition α -helix and specific base-pairs in the major groove.

Although structures of the free DNA sites have not been determined, Joshi *et al.*, attribute the key differences in minor-groove interactions to intrinsic sequence-dependent structural variations. This hypothesis is supported by the respective bound-state structures and DNA simulations. In the paralog-specific DNA site (as expected for an A-T-rich sequence) the minor groove is narrower on either side of the central T-A step (boldface above); in contrast the consensus DNA site exhibits narrowing of the minor groove *only to the left of the T-A step* (5' to the bases in bold). Whereas in each complex an unrelated Arg side chain binds to the shared 5'-narrowed minor groove, the conserved Arg-His element binds only to the 3'-narrowed minor groove specific to the paralog site. Local differences in electrostatic potential, predicted to be a consequence of respective minor-groove widths is proposed to underlie such DNA sequence discrimination by the Scr-Exd and in turn its specific role in development. This model envisages that unlike hydrogen-bond-based recognition code in the major groove, minor-groove recognition by the N-terminal arm of the homeodomain employs indirect readout of a preformed sequence-dependent DNA structure. However appealing, this model remains

speculative in the absence of the structures of the respective free DNA sites. Nonetheless, the present results support the model of Joshi *et al.*, as the minor-groove width variations observed in the two Scr-Exd-DNA complexes are recapitulated in the structure of a *free* DNA element. In *dsx^A* the smaller minor groove width occurs on either side of the central T-G step, analogous to the key T-A step in the Scr-Exd-DNA complex. The sequence-dependent pattern of minor-groove widths in *dsx^A* is illustrated in Figures 8 and 9, wherein the width is smallest at the central T-G step of the duplex (Figure 8A) with gradual increase on either side (Figure 8B).

DNA bending

The structure of the *dsx^A* site exhibits slight DNA bending. Whereas the 5'-TACAATGTTGCA-3' region is nearly straight ($\sim 4^\circ$ bend), the two base-pair steps at each end have positive roll angles, suggesting a roll-bend toward the major groove. Occurrence of minor kinks ($\sim 8^\circ$) at T5-A6 (underlined in 5'-TACAATGTTGCA-3') and T23-G24 (underlined in complementary strand 5'-TGCAACATTGTA-3') flanking the terminal two base pairs at each end is in accord with the general observation that such kinking is associated with pyrimidine-purine steps^{43,44} with little free-energy penalty.⁴⁵ Anisotropic bending typically results from positive roll at C-A/T-G, T-A or less frequently C-G steps, in a direction that compresses the major groove.²³ The calcium ions bound in the minor groove do not induce bending as noted previously in a decamer structure.¹² Confinement of DNA bending to the ends of the *dsx^A* site raises the caveat that interpretation of crystal structures of nucleic acids requires consideration of end effects. Although in the present case such effects may be attenuated by formation of a pseudo-continuous double helix in the crystal lattice, a periodic discontinuity necessarily occurs in the DNA backbone.

To minimize possible effects of this discontinuity, it is instructive to focus on the central 8-base-pair segment (5'-ACAATGTT-3' and complement). As noted above, this segment is separated by four base pairs at each side from the end of the DNA duplex, an interval proposed to insulate the interior segment from transmitted conformational perturbations.⁴¹ In the central 8-base-pair segment the mean twist (10.7 base pairs/turn) is close to 10.6 base pairs/turn found in solution for bulk B-DNA,⁴⁶ suggesting that this segment is not significantly affected by crystal packing forces. Sequence-dependent variations in its structure and dynamics may therefore illuminate general features of B-DNA. The inner sub-segment 5'-AATGTT-3' (and complement) possesses the structural hallmarks of a straight A-T-tract without a T-A step. To our knowledge, this is the first example of a sequence mimicking a straight A-T-tract despite the presence of an interrupting T-G step, whose intrinsic flexibility is close to or even larger than that of a T-A step.⁴⁷ For this reason we have chosen to discuss aspects of the structure of the central segment in detail.

Structure and Dynamics of a Pseudo-A-T tract

Within the central segment, the sequence 5'-AATGTT-3' (and its complement) is associated with narrowing of the minor groove, a well-organized spine of hydration, and high propeller twist (mean = -15°); these features appear intimately linked. Absent are bifurcated hydrogen bonds adjoining the major groove, features that are by contrast prominent in A-tracts containing four or more consecutive adenines.³⁴ Extent of propeller twist at a specific base pair depends on its sequence context as illustrated by base pairs C7-G32 and G11-C28: these respective sites exhibit low (-3°) and high (-14°) propeller twist. Due to such sequence-specific features, the A6-C7 step differs in geometry, groove width, and hydration from the neighboring A8-A9-T10-G11-T12-T13 sub-segment and so defines a boundary between structural units. Further, in the T-G step (T10-G11/C28-A29) the deoxyribose moieties of T10, G11 and C28 – as well as the phosphate group of A29 – exhibit poor electron density with higher thermal *B*-factors (Figure 3). The *B*-factors and electron density are correlated, wherein a high atomic displacement parameter for a specific atom or a group of atoms is generally associated with

poor or no electron density. Such anomalous *B*-factors suggest a malleable element within an otherwise rigid pseudo-A-T-tract. The flexibility of a T-G step is in agreement with previous empirical and theoretical calculations.⁴⁷

The 6-base-pair sub-segment A8-A9-T10-G11-T12-T13 exhibits additional structural features distinct from the rest of the structure. The innermost step (T10-G11) has a unique geometry (Table 2) with high twist (37.6°) and the largest values of rise (3.6 Å) and roll (8.5°). Base pairs T10-A29 and G11-C28 display high buckling (7°, -13°), propeller twist (-16°, -14°), and opening (4°, 4°). In addition, a pseudo-two-fold symmetry axis orthogonal to the helical axis passes through the middle of this step. The three base-pair steps flanking the rigid core resemble alternating B-DNA^{24,31} with a high-low-high twist pattern at T5-A6 (41.5°), A6-C7 (21.8°), C7-A8 (47.9°); T13-G14 (49.1°), G14-C15 (21.5°), and C15-A16 (42.0°) steps. This pattern suggests that base stacking at pyrimidine-purine steps is weak relative to that at purine-pyrimidine steps. In the alternating B-DNA model of poly(dA-dT), twists at A-T and T-A steps alternate between 31° and 40° as observed at A-T and T-A steps in the crystal structure of d (CGCATATATGCG).³¹ In the Klug alternating B-DNA model torsion angles about P-O3' bonds (ζ) are significantly different for the two base-pair steps ($\sim -70^\circ$ for A-T and $\sim -125^\circ$ for T-A). Sugar-phosphate geometries in the central segment of *dsx*^A assume a standard B_I conformation at purine-pyrimidine steps ($\zeta \sim -100^\circ$), whereas B_{II} conformations are observed at pyrimidine-purine steps ($\zeta \sim 165^\circ$) in the alternating B-DNA region (C7-A8, T13-G14, C15-A16; C25-A26, T31-G32, and T33-A34). Further, torsion angles about O5'-C5' bonds (β) differ considerably from the normal value ($175^\circ \pm 6^\circ$) at bases G14, A16, G32, and A34 with an average value of $141^\circ \pm 1^\circ$.

The distinctive features of the central 5'-ATG-3' element in *dsx*^A may be considered in the context of the existing crystallographic database of DNA structures and protein-DNA complexes.⁴⁸ An empirical correlation between DNA sequence and helical parameters was obtained by comparative analysis of dinucleotide steps.⁴⁹ These studies proposed a kinematic classification of dinucleotide steps into two classes, rigid or loose; the latter may be flexible or bistable. This classification is only approximate however, as effects of flanking DNA sequences were omitted. Although the intrinsic properties of dinucleotide steps are indeed distinct, it is also necessary to include effects of flanking steps as suggested by Dickerson and co-workers.⁴¹ We have thus compared the structure of the 5'-ATG-3' element in *dsx*^A to such elements in an extensive collection of crystal structures.⁴⁸ Three key parameters (rise, roll, and twist) are in general smaller at A-T steps than at T-G steps. In the majority of structures the central T-A base pair (bold) among 5'-ATG-3' elements exhibits greater buckle and a higher opening angle than do the flanking base pairs. The majority of 5'-ATG-3' elements in the data base (38 out of 65 examples) nonetheless exhibit a mixture of behaviors with no predominant pattern, perhaps due to overarching effects of crystal packing and specific interactions with binding proteins. Propeller twists also lack a definitive pattern (Supporting Information).

Minor groove spine of hydration

Solvation of DNA has long been known to influence its conformation.^{50,51} Specific bound water molecules can also modulate its interaction with proteins and other ligands.⁵ The minor- and major grooves of B-DNA display distinctive patterns of hydration.^{7,52} In the present study the structure of bound water molecules in the minor groove exhibits a pseudo-two-fold symmetry about the central T-G step. Despite this intervening T-G step, the central segment retains a canonical spine of hydration (Figure 6) as previously associated with the narrowed AATT minor groove of B-DNA.^{7,53,54} Incorporation of a G-C base pair within an A-T-rich core is thus well tolerated in this specific sequence context. The importance of context is illustrated by the contrasting effect of the C7-G32 base pair flanking the core pseudo-A-T-tract (underlined 5'-ACAATGTT-3' and complement). The spine of hydration begins at A8 (bold

above) and not from A6 (italics) because the intervening base (C7) is incompatible with the spine structure. This interruption suggests that at least two consecutive adenines are required to nucleate the spine. The particular structure of the hydration spine may underlie the empirical association between narrowing of the minor groove and high propeller twist that has been generally observed in crystal structures of unbent A-T-rich DNA duplexes.⁵³ We speculate that the symmetrical and highly organized hydration pattern in the minor groove of *dsx*^A may have a significant role “indirect readout” by binding proteins.

Calcium binding

Each end of the central spine contains two hydrated calcium ions (Figure 6) as in R3 DDD structures.^{7,54} In the case of orthorhombic form of DDD structures, Mg^{2+} ions border the spine in similar positions.⁵⁵⁻⁵⁷ Here, however, there are two additional calcium ions at the duplex termini. The minor groove is thus continuously occupied in succession by hydrated calcium ions, a canonical spine of hydration, and other ordered water molecules (Figure 6). The calcium ions in *dsx*^A are surrounded by water molecules and/or phosphate oxygens with an average Ca^{2+} -O distance of 2.4 Å in accord with previous studies.⁵⁸ Binding of calcium ions in the minor groove is dependent on groove width, which in turn is dependent on sequence. Crystals in the present study belong to R3 space group, like DDD crystals grown in the presence of calcium ions (R3 crystal form^{7,54}). Stacking of duplexes in *dsx*^A crystals and such R3 DDD crystals nonetheless differ in relative inter-helical twists (-25.5° versus -70.5°). Although the present structure is uniquely stabilized by bridging Ca^{2+} ions at opposing ends successive duplexes, both *dsx*^A and R3 DDD crystals exhibit analogous calcium-bridged interactions between neighboring columns related by a crystallographic three-fold symmetry. These structures demonstrate that divalent metal ions found at the interface of duplexes can influence their mode of packing by specific interactions, *i.e.*, beyond the general electrostatic screening of the anionic DNA backbone.

A Model for DNA Dynamics?

The *B*-factors of each nucleotide in a DNA structure, considered by class of atom, reflect the extent of static- and/or dynamic disorder. The *B*-factors in *dsx*^A (Figure 3) gradually rise from the ends of the duplex to a maximum within the central sub-segment: the innermost T10-A29 and G11-C28 base pairs display the highest *B* values ($22 \text{ \AA}^2$; mean $=15 \pm 4 \text{ \AA}^2$). In cases of conformational variability or discrete disorder, ensembles of structures, which are simultaneously refined against the X-ray data, may serve as a more appropriate description of the observed data. A noteworthy application of multi-conformer crystallographic refinement enabled the simulation of flexible loops in a protein.⁵⁹ In this protocol multiple copies of the structure – with partial occupancy assigned to each – were simultaneously refined with a fixed and uniform overall *B*-factor. We have attempted by analogy to simulate the portions of the DNA duplex proposed to exhibit intrinsic flexibility by utilizing molecular dynamics⁶⁰ as a visualization tool (Supplemental Information). The resulting ensemble of models (Supplemental Information) may be tested through NMR studies of a corresponding DNA duplex in solution. We envision that such heuristic dynamic models may be of value by making quantitative predictions regarding the range of conformers occupied in solution at each base step.

Observation of sequence-specific variation in *B*-factors and quality of the electron density in A9-T10-G11 region motivated us to obtain a comparative analysis of sequence-specific helical parameters in the three-base-pair sub-site A-T-G/C-A-T in previously described DNA structures. In this analysis, crystal structures of free DNA fragments involving only unmodified Watson-Crick base pairs or in complex with protein determined at 2 Å or higher resolution and possesses A-T-G/C-A-T triplet in a non distorted or kinked DNA segment were considered. From a set of 4 free DNA and 30 DNA-protein complexes, we thus found 65 such triplets and

their local helical parameters were analyzed using program 3DNA.³⁹ Base-pair and step parameters associated with 65 A-T-G subsites in a variety of sequences were examined with the aim of uncovering any underlying correlations (Supporting Information). This analysis suggests general trends. As in the present structure, the rise at A-T step ($3.2 \text{ \AA} \pm 0.1 \text{ \AA}$) is smaller than that found at T-G step ($3.4 \text{ \AA} \pm 0.2 \text{ \AA}$) in 77 % of the cases (50/65). Likewise the roll value is smaller at A-T step ($1.8^\circ \pm 4.0^\circ$) compared to adjacent T-G step ($4.3^\circ \pm 5.0^\circ$) in most cases (44/65). Similarly the helical twist at A-T step ($29.0^\circ \pm 4.1^\circ$) is smaller than the consecutive T-G step ($37.1^\circ \pm 6.2^\circ$) in 49 cases. It is important to note that large root-mean-square-deviation values found for rolls and twists imply wide variation, nonetheless the successive steps possess a distinctive pattern (Supporting Information). Similarly, the occurrence of disordered deoxyribose of C28 is in line with the recent DNA simulation studies that suggests unusual conformational propensity for sugars attached to cytosine.⁶¹ To our knowledge, the present structure provides the first high-resolution crystallographic evidence of base-pair-specific disorder within an otherwise well-organized DNA structure.

A proposal for induced fit in an SRY-DNA complex

The SRY HMG box recognizes a class of AT-rich sites with core consensus ACAA(T/A) and hence binds with high affinity to the central segment of *dsx*^A (5'-ACACTACAATGTTGCAAT-3' and complement; core consensus sequence in bold). Fluorescence resonance energy transfer studies suggest that the SRY domain bends *dsx*^A to an extent similar to that observed in the NMR structure of a specific complex with the DNA site⁶² (designated the Mullerian Inhibiting Substance (*MIS*) element; 5'-CCTGCACAAACCC-3' and complement; L. Zhou and M. A. W., unpublished results). Because to our knowledge there is no structure of the free *MIS* DNA site (or other SRY-related DNA target site), we may utilize the present structure as a baseline for analysis of induced fit. Comparison between the free structure of *dsx*^A and the bound structure of the *MIS* element is restricted to the central 5'-ACAA-3' subsite shared between the two sequences. The marked changes in structural parameters within this subsite are summarized in Table 3. Of particular interest are the changes in base-pair roll and propeller twist at the site of partial intercalation by the "cantilever" side chain of the SRY HMG box (I13 in the HMG-box consensus).⁶² The successive roll angles are large and positive (feature of a bent DNA) in the *MIS* element whereas such angles are smaller and exhibit alternating positive and negative values in the unbound *dsx*^A structure. This striking change highlights the extent of induced fit. Comparable differences are observed throughout the subsite (Table 3), implying global reorganization of DNA structure upon binding of SRY.

Implications for *dsx*^A-DM complex

Doublesex contains a DM domain, a conserved motif of minor-groove DNA recognition. This motif binds cooperatively as a homodimer in the minor groove of *dsx*^A.^{21,63} Unlike other minor-groove-binding proteins (such as the TATA-binding protein TBP and HMG boxes,⁶⁴ the DM domain does not induce sharp DNA bending. This unusual minor groove binding without bending resembles the T domain-DNA complex of the *Brachyury* transcription factor.²⁷ Molecular recognition of the DNA minor groove is in part based on structural deformations rather than its registry of hydrogen-bond donors and acceptors.⁶⁴ Although the structure of the DSX-*dsx*^A complex has not been determined, we predict that discrimination of A·T *versus* T·A and C·G *versus* G·C base pairs will exploit some of the features observed here in the free DNA: the malleability of specific sites, including the central T-G step, and sequence-dependent organization of bound water molecules. Indeed, chemogenetic experiments, employing nonstandard DNA analogs and modified DM domains, have demonstrated that the protein contacts *dsx*^A immediately flanking the T-G and complementary C-A steps.⁶⁵ The pseudo-two-fold symmetry of *dsx*^A duplex (itself not a palindrome) may foreshadow its mode of binding to DSX protein homodimers. Although the co-crystal structure of a DM-DNA complex

has to date not been determined, the biological importance of sequence-dependent differences in DNA structure and dynamics at AT-rich sub-sites was recently suggested by analysis of homeodomain-DNA complexes.⁴² Interpretation of the Hox-DNA complex by Joshi et al.,⁴² was limited by the absence of structures for the free DNA sites. Whereas simulations suggested that structural features of the protein-DNA complexes are likely to the DNA target sequences, further progress will require joint analyses of free and bound DNA structures. Such shared structural principles may underlie the functional inter-relationships between homeotic genes and DM genes in the developmental patterning of metazoans.⁶⁶

Materials and Methods

DNA synthesis, crystallization and data collection

Oligonucleotides d(5'-ACACTACAATGTTGCAAT-3') and d(5'-GTATTGCAACATTGTAGT-3') were purchased from Operon Biotechnologies Inc. (Huntsville, Alabama) with the 5'-trityl on. Samples were purified by reverse-phase high-performance liquid chromatography on a C4 column; elution was effected by an acetonitrile gradient, followed by mild acetic-acid treatment to remove the trityl group, and re-chromatography on the same column. The complementary single strands were then annealed in a 1:1 ratio to form a DNA duplex. A similar protocol was followed to obtain duplexes with 5-bromo-deoxyuridine (BrU) containing derivatives. BrU substitutions were made at two positions (T18 and T23; top panel of Figure 1) to enable multiple anomalous dispersion (MAD) phasing.

Crystals were grown in hanging drops containing 0.25 mM DNA duplex, 5 mM CaCl₂, 0.2 M KCl, 5% (w/v) polyethylene glycol-10000 (PEG10K), 5% (v/v) methyl-2,4-pentane-diol (MPD), and 50 mM Tris-HCl (pH 8.0). These drops were equilibrated against a reservoir of 10% PEG10K, 10% MPD, and 100 mM Bicine-HCl buffer (pH 8.0) at 4 °C. Crystals suitable for X-ray diffraction ($0.2 \times 0.25 \times 0.3$ mm³) grew in about four weeks. Prior to data collection, crystals were flash-cooled to 100 °K using 30% (v/v) PEG400 as cryoprotectant. The crystals belong to space group R3 with unit cell dimensions of $a = b = 38.74$ Å, $c = 161.33$ Å and $\gamma = 120.0^\circ$. Assuming a standard value for the partial specific volume of B-DNA,⁶⁷ the crystal was calculated to have a solvent content of 54% (Matthews coefficient $V_m = 2.19$ Å³/Da), similar to that of other DNA crystals.⁶⁸ This crystal form contains one DNA duplex in the asymmetric unit. MAD data from a single bromine-derivatized crystal were collected at wavelengths based on its fluorescence spectrum. Native data were collected to 1.6 Å resolution (99% complete; $R_{\text{sym}} = 0.035$) at MacCHESS beam line A1. A summary of crystallographic data is provided in Table 4. Data were processed with programs DENZO and SCALEPACK.⁶⁹

Structure determination and refinement

The DNA structure was determined utilizing MAD phases from the BrU analog using programs SOLVE and RESOLVE with inclusion of solvent flattening.⁷⁰ Nucleic-acid residues were built into the electron density map calculated at 2.2 Å resolution. Table 4 contains a summary of MAD phasing statistics. Refinement was initiated using native data with program CNS;⁷¹ model adjustments were performed using O.⁷² In each round SA torsion-angle molecular-dynamics refinement was performed followed by individual *B*-factor refinement as implemented in CNS. 10% of the unique data were set aside to calculate R_{free} . SA refinements were interspersed with manual model rebuilding and gradual inclusion of higher-resolution data. The first two residues at the 5'-end of each strand were not seen in electron density maps and are hence assumed to be disordered. The R-factor converged at 35% for an intermediate model containing residues 3-18 and 21-36 for data between 20-1.6 Å. At this stage water molecules were included automatically using the water-picking utility within CNS, and the correctness of their positioning was evaluated by visualizing difference electron-density maps

on a graphics display. A water molecule was assigned to a peak that was observed at a potential hydrogen-bonding distance from a DNA atom or another water molecule. In total 197 water molecules were thus located in eight cycles of successive refinement. Peaks corresponding to W62, W92, W145, and W221 assigned as water molecules that are located near the missing overhung 5'-end residues may correspond to partially occupied O or N atoms of the mobile residues (G19 and T20). The model was further optimized by using omit maps, computed by omitting two nucleotide residues at a time. In final stages of the refinement several omit maps were computed by omitting 10 water molecules at a time. In these maps six sharp peaks were observed at greater than 20 σ and another six peaks at about 12 σ , with coordination geometries in each case compatible with calcium ions.⁵⁷ Further refinement with inclusion of 12 calcium ions resulted in a final R-factor of 22.1% for data between 8-1.6 Å resolution. Representative sections of the DNA structure and its associated electron density are shown in Figure 2. A summary of refinement statistics is presented in Table 4.

Coordinates

Atomic coordinates and structure factor files are deposited with the Protein Data Bank (accession code 3BSE).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to Prof. Bruce Baker (Stanford University) for advice and encouragement pursuant to studies of Doublesex. We thank U. Narendra for purifying oligonucleotides and crystallization trials, M. Szebenyi and the staff at Cornell High Energy Synchrotron Source (MacCHESS) facility for help in data collection and processing, S. B. Kent for synthesis of DSX DM polypeptides, A. R. Srinivasan and W. K. Olson (Rutgers University) for advice on the manuscript. NN was supported in part by an institutional pilot grant from the Ohio Chapter of the American Cancer Society (IRG-91-022-09) and the Ohio Cancer Research Associates. This work was supported in part by a subcontract to M.A.W. from the National Institutes of Health (R01 GM037731 to B. Baker). This article represents a contribution from the Cleveland Center for Structural Biology.

References

1. Rhodes D, Schwabe JW, Chapman L, Fairall L. Towards an understanding of protein-DNA recognition. *Philos Trans R Soc Lond B Biol Sci* 1996;351:501–509. [PubMed: 8735272]
2. Travers AA. The structural basis of DNA flexibility. *Philos Transact A Math Phys Eng Sci* 2004;362:1423–1438. [PubMed: 15306459]
3. von Hippel PH, Rees WA, Rippe K, Wilson KS. Specificity mechanisms in the control of transcription. *Biophys Chem* 1996;59:231–246. [PubMed: 8672714]
4. Seeman NC, Rosenberg JM, Rich A. Sequence-specific recognition of double helical nucleic acids by proteins. *Proc Natl Acad Sci U S A* 1976;73:804–808. [PubMed: 1062791]
5. Shakked Z, Guzikovich-Guerstein G, Frolov F, Rabinovich D, Joachimiak A, Sigler PB. Determinants of repressor/operator recognition from the structure of the trp operator binding site. *Nature* 1994;368:469–473. [PubMed: 8133895]
6. Sines CC, McFail-Isom L, Howerton SB, VanDerveer D, Williams LD. Cations mediate B-DNA conformational heterogeneity. *J Am Chem Soc* 2000;122:11048–11056.
7. Minasov G, Tereshko V, Egli M. Atomic-resolution crystal structures of B-DNA reveal specific influences of divalent metal ions on conformation and packing. *J Mol Biol* 1999;291:83–99. [PubMed: 10438608]
8. Kim Y, Geiger JH, Hahn S, Sigler PB. Crystal structure of a yeast TBP/TATA-box complex. *Nature* 1993;365:512–520. [PubMed: 8413604]

9. Kim JL, Nikolov DB, Burley SK. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* 1993;365:520–527. [PubMed: 8413605]
10. Travers A. Recognition of distorted DNA structures by HMG domains. *Curr Opin Struct Biol* 2000;10:102–109. [PubMed: 10679469]
11. Weiss MA. Floppy SOX: mutual induced fit in HMG (high-mobility group) box-DNA recognition. *Mol Endocrinol* 2001;15:353–362. [PubMed: 11222737]
12. Chiu TK, Dickerson RE. 1 Å crystal structures of B-DNA reveal sequence-specific binding and groove-specific bending of DNA by magnesium and calcium. *J Mol Biol* 2000;301:915–945. [PubMed: 10966796] *J Mol Biol* 303:111. Erratum in
13. Koo HS, Wu HM, Crothers DM. DNA bending at adenine thymine tracts. *Nature* 1986;320:501–506. [PubMed: 3960133]
14. Rozenberg H, Rabinovich D, Frolov F, Hegde RS, Shakked Z. Structural code for DNA recognition revealed in crystal structures of papillomavirus E2-DNA targets. *Proc Natl Acad Sci U S A* 1998;95:15194–15199. [PubMed: 9860945]
15. Hizver J, Rozenberg H, Frolov F, Rabinovich D, Shakked Z. DNA bending by an adenine--thymine tract and its role in gene regulation. *Proc Natl Acad Sci U S A* 2001;98:8490–8495. [PubMed: 11438706]
16. Burtis KC, Baker BS. *Drosophila* doublesex gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. *Cell* 1989;56:997–1010. [PubMed: 2493994]
17. Grzeskowiak K. Sequence-dependent structural variation in B-DNA. *Chem Biol* 1996;3:785–790. [PubMed: 8939695]
18. Withka JM, Swaminathan S, Srinivasan J, Beveridge DL, Bolton PH. Toward a dynamical structure of DNA: comparison of theoretical and experimental NOE intensities. *Science* 1992;255:597–599. [PubMed: 1736362]
19. Boisbouvier J, Delaglio F, Bax A. Direct observation of dipolar couplings between distant protons in weakly aligned nucleic acids. *Proc Natl Acad Sci U S A* 2003;100:11333–11338. [PubMed: 12972645]
20. An W, Wensink PC. Three protein binding sites form an enhancer that regulates sex- and fat body-specific transcription of *Drosophila* yolk protein genes. *EMBO J* 1995;14:1221–1230. [PubMed: 7720712]
21. Zhu L, Wilken J, Phillips NB, Narendra U, Chan G, Stratton SM, Kent SB, Weiss MA. Sexual dimorphism in diverse metazoans is regulated by a novel class of intertwined zinc fingers. *Genes Dev* 2000;14:1750–1764. [PubMed: 10898790]
22. Kim S, Namekawa SH, Niswander LM, Ward JO, Lee JT, Bardwell VJ, Zarkower D. A mammal-specific Doublesex homolog associates with male sex chromatin and is required for male meiosis. *PLoS Genet* 2007;3:e62. [PubMed: 17447844]
23. Dickerson RE. DNA bending: the prevalence of kinkiness and the virtues of normality. *Nucleic Acids Res* 1998;26:1906–1926. [PubMed: 9518483]
24. Klug A, Jack A, Viswamitra MA, Kennard O, Shakked Z, Steitz TA. A hypothesis on a specific sequence-dependent conformation of DNA and its relation to the binding of the lac-repressor protein. *J Mol Biol* 1979;131:669–680. [PubMed: 513130]
25. Rhodes D, Klug A. Sequence-dependent helical periodicity of DNA. *Nature* 1981;292:378–380. [PubMed: 6265794]
26. Wang JC. Helical repeat of DNA in solution. *Proc Natl Acad Sci U S A* 1979;76:200–203. [PubMed: 284332]
27. Muller CW, Herrmann BG. Crystallographic structure of the T domain-DNA complex of the Brachyury transcription factor. *Nature* 1997;389:884–888. [PubMed: 9349824]
28. Drew HR, Samson S, Dickerson RE. Structure of a B-DNA dodecamer at 16 K. *Proc Natl Acad Sci U S A* 1982;79:4040–4044. [PubMed: 6955789]
29. Kielkopf CL, Ding S, Kuhn P, Rees DC. Conformational flexibility of B-DNA at 0.74 Å resolution: d(CCAGTACTGG)₂. *J Mol Biol* 2000;296:787–801. [PubMed: 10677281]
30. Chandrasekaran R, Arnott S. The structure of B-DNA in oriented fibers. *J Biomol Struct Dyn* 1996;13:1015–1027. [PubMed: 8832384]

31. Yoon C, Prive GG, Goodsell DS, Dickerson RE. Structure of an alternating-B DNA helix and its relationship to A-tract DNA. *Proc Natl Acad Sci U S A* 1988;85:6332–6336. [PubMed: 3413099]
32. Coll M, Frederick CA, Wang AH, Rich A. A bifurcated hydrogen-bonded conformation in the d(A.T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin. *Proc Natl Acad Sci U S A* 1987;84:8385–8389. [PubMed: 3479798]
33. DiGabriele AD, Steitz TA. A DNA dodecamer containing an adenine tract crystallizes in a unique lattice and exhibits a new bend. *J Mol Biol* 1993;231:1024–1039. [PubMed: 8515463]
34. Nelson HC, Finch JT, Luisi BF, Klug A. The structure of an oligo(dA).oligo(dT) tract and its biological implications. *Nature* 1987;330:221–226. [PubMed: 3670410]
35. Johansson E, Parkinson G, Neidle S. A new crystal form for the dodecamer C-G-C-G-A-A-T-T-C-G-C-G: symmetry effects on sequence-dependent DNA structure. *J Mol Biol* 2000;300:551–561. [PubMed: 10884351]
36. Lavery R, Sklenar H. The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J Biomol Struct Dyn* 1988;6:63–91. [PubMed: 2482765]
37. Dickerson, RE.; Kopka, ML.; Drew, HR. Structural correlations in B-DNA. In: Srinivasan, R.; Sarma, RH., editors. *Conformation in Biology*. Adenine Press; New York, NY: 1982. p. 227-257.
38. Grzeskowiak K, Yanagi K, Prive GG, Dickerson RE. The structure of B-helical C-G-A-T-C-G-A-T-C-G and comparison with C-C-A-A-C-G-T-T-G-G. The effect of base pair reversals. *J Biol Chem* 1991;266:8861–8883. [PubMed: 2026600]
39. Lu XJ, Olson WK. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res* 2003;31:5108–5121. [PubMed: 12930962]
40. Straus D, Schlick T. A-Tract bending: insights into experimental structures by computational models. *J Mol Biol* 2000;301:643–663. [PubMed: 10966775]
41. Yanagi K, Prive GG, Dickerson RE. Analysis of local helix geometry in three B-DNA decamers and eight dodecamers. *J Mol Biol* 1991;217:201–214. [PubMed: 1988678]
42. Joshi R, Passner JM, Rohs R, Jain R, Sosinsky A, Crickmore MA, Jacob V, Aggarwal AK, Honig B, Mann RS. Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell* 2007;131:530–543. [PubMed: 17981120]
43. Dickerson RE, Chiu TK. Helix bending as a factor in protein/DNA recognition. *Biopolymers* 1997;44:361–403. [PubMed: 9782776]
44. McNamara PT, Bolshoy A, Trifonov EN, Harrington RE. Sequence-dependent kinks induced in curved DNA. *J Biomol Struct Dyn* 1990;8:529–538. [PubMed: 2100517]
45. Protozanova E, Yakovchuk P, Frank-Kamenetskii MD. Stacked-unstacked equilibrium at the nick site of DNA. *J Mol Biol* 2004;342:775–785. [PubMed: 15342236]
46. Rhodes D, Klug A. Sequence-dependent helical periodicity of DNA. *Nature* 1981;292:378–380. [PubMed: 6265794]
47. Packer MJ, Dauncey MP, Hunter CA. Sequence-dependent DNA structure: dinucleotide conformational maps. *J Mol Biol* 2000;295:71–83. [PubMed: 10623509]
48. The PDB accession codes for the crystal structures used in this analysis are 1d8x, 1du0, 1e3o, 1gtw, 1l3l, 1lat, 1nfx, 1nk7, 1owf, 1puf, 1rh6, 1rxw, 1tv9, 1ua1, 1wqy, 1xc9, 1xo0, 1yo5, 2a07, 2ac0, 2ahi, 2asd, 2bam, 2bdp, 2euz, 2evz, 2fjw, 2fmp, 2hdd, 2nll, 3bdp, 4bdp, 126d, and 334d.
49. El Hassan MA, Calladine CR. Conformational characteristics of DNA: empirical classifications and a hypothesis for the conformational behaviour of dinucleotide steps. *Philos Trans R Soc London, Ser A* 1997;355:43–100.
50. Franklin RE, Gosling RG. Molecular configuration in sodium thymonucleate. *Nature* 1953;171:740–741. [PubMed: 13054694]
51. Arnott S, Chandrasekaran R, Hall IH, Puigjaner LC, Walker JK, Wang M. DNA secondary structures: helices, wrinkles, and junctions. *Cold Spring Harb Symp Quant Biol* 1983;47:53–65. [PubMed: 6574857]
52. Tereshko V, Minasov G, Egli M. A “Hydrat-Ion” spine in a B-DNA minor groove. *J Am Chem Soc* 1999;121:3590–3595.
53. Kopka ML, Fratini AV, Drew HR, Dickerson RE. Ordered water structure around a B-DNA dodecamer. A quantitative study. *J Mol Biol* 1983;163:129–146. [PubMed: 6834428]

54. Liu J, Subirana JA. Structure of d(CGCGAATTCGCG) in the presence of Ca²⁺ ions. *J Biol Chem* 1999;274:24749–24752. [PubMed: 10455145]
55. Soler-Lopez M, Malinina L, Liu J, Huynh-Dinh T, Subirana JA. Water and ions in a high resolution structure of B-DNA. *J Biol Chem* 1999;274:23683–23686. [PubMed: 10446123]
56. Tereshko V, Minasov G, Egli M. The Dickerson-Drew B-DNA dodecamer revisited at atomic resolution. *J Am Chem Soc* 1999;121:470–471.
57. Chiu TK, Kaczor-Grzeskowiak M, Dickerson RE. Absence of minor groove monovalent cations in the crosslinked dodecamer C-G-C-G-A-A-T-T-C-G-C-G. *J Mol Biol* 1999;292:589–608. [PubMed: 10497024]
58. Katz AK, Glusker JP, Beebe SA, Bock CW. Calcium ion coordination: A comparison with that of beryllium, magnesium, and zinc. *J Am Chem Soc* 1996;118:5752–5763.
59. Gill HS, Pfluegl GM, Eisenberg D. Multicopy crystallographic refinement of a relaxed glutamine synthetase from *Mycobacterium tuberculosis* highlights flexible loops in the enzymatic mechanism and its regulation. *Biochemistry* 2002;41:9863–9872. [PubMed: 12146952]
60. Kuriyan J, Osapay K, Burley SK, Brunger AT, Hendrickson WA, Karplus M. Exploration of disorder in protein structures by X-ray restrained molecular dynamics. *Proteins* 1991;10:340–358. [PubMed: 1946343]
61. Olson, WK.; Colasanti, AV.; Li, Y.; Ge, Wei; Zheng, G.; Zhurkin, VB. DNA simulation benchmarks as revealed by x-ray structures. In: Sponer, J.; Lankas, F., editors. *Computational studies of RNA and DNA*. Springer; 2006. p. 235-257.
62. Murphy EC, Zhurkin VB, Louis JM, Cornilescu G, Clore GM. Structural basis for SRY-dependent 46-X,Y sex reversal: modulation of DNA bending by a naturally occurring point mutation. *J Mol Biol* 2001;312:481–499. [PubMed: 11563911]
63. Narendra U, Zhu L, Li B, Wilken J, Weiss MA. Sex-specific gene regulation: the doublesex DM motif is a bipartite DNA-binding domain. *J Biol Chem* 2002;277:43463–43473. [PubMed: 12198117]
64. Bewley CA, Gronenborn AM, Clore GM. Minor groove-binding architectural proteins: structure, function, and DNA recognition. *Annu Rev Biophys Biomol Struct* 1998;27:105–131. [PubMed: 9646864]
65. Zhang W, Li B, Singh R, Narendra U, Zhu L, Weiss MA. Regulation of sexual dimorphism: mutational and chemogenetic analysis of the doublesex DM domain. *Mol Cell Biol* 2006;26:535–547. [PubMed: 16382145]
66. Chen EH, Christiansen AE, Baker BS. Allocation and specification of the genital disc precursor cells in *Drosophila*. *Dev Biol* 2005;281:270–285. [PubMed: 15893978]
67. Schildkraut CL, Marmur J, Doty P. Determination of the base composition of deoxyribonucleic acid from its buoyant density in CsCl. *J Mol Biol* 1962;4:430–443. [PubMed: 14498379]
68. Kennard O, Hunter WN. Oligonucleotide structure: a decade of results from single crystal X-ray diffraction studies. *Q Rev Biophys* 1989;22:327–379. [PubMed: 2695962]
69. Otwinowski Z, Minor W. Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol* 1997;276:307–326.
70. Terwilliger TC. SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol* 2003;374:22–37. [PubMed: 14696367]
71. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for Macromolecular Structure determination. *Acta Crystallogr, Sect D* 1998;54:905–921. [PubMed: 9757107]
72. Jones TA. Interactive electron-density map interpretation: from INTER to O. *Acta Crystallogr, Sect D* 2004;60:2115–2125. [PubMed: 15572764]

Abbreviations

BrU	5-bromo-deoxyuridine
Ca	calcium

DDD	Dickerson-Drew dodecamer
DM	N-terminal DNA-binding domain of DSX
DNA	deoxyribose nucleic acid
DSX	doublesex
HMG	high mobility group
MAD	multiple anomalous dispersion
MPD	methyl-2,4-pentane-diol
PDB	Protein Data Bank
PEG	polyethylene glycol
RMSD	root-mean-square deviation
SA	simulated annealing
S.D	standard deviation
TBP	TATA-box binding protein.

Nucleic-acid residues are designated by standard single-letter codes.

Figure 1.

Sequence and structure of the *dsx^A* site. Top panel shows the DNA sequence with two-residue overhangs at the 5'-ends with the numbering scheme. The sequence within the box represents the consensus binding site of the DSX protein. Bottom panel depicts a stereo view of a stick representation of the overall structure. The residues in the overhang on both the strands were not seen in the electron density map. The split-bond coloring scheme is based on the atom colors with carbon atoms in green, nitrogen atoms in blue, oxygen atoms in red, and phosphorous atoms in yellow.

Figure 2.

Electron density map of 16 base pairs. A $2F_o - F_c$ map computed using the final refined model is shown contoured at 1.0 σ . The dotted lines indicate hydrogen bonds. The stick representation of nucleic acid residues bears the same coloring scheme as in Figure 1. (A) – (P) shows electron density for individual base pairs beginning from A3·T36, C4·G35, T5·A34, A6·T33, C7·G32, A8·T31, A9·T30, T10·A29, G11·C28, T12·A27, T13·A26, G14·C25, C15·G24, A16·T23, A17·T22, and T18·A21. Some base pairs display excellent density for all atoms (for example, see panels A – D). In panel H (T10·A29), the deoxyribose of T10 exhibits poor density. In panel I (G11·C28) the deoxyribose rings of G11 and C28 possess weak or no density suggesting large mobility.

Figure 3. *B*-factor plot. The group temperature factors (Å) of bases (purple), deoxyribose sugars (green), phosphates (red), and nucleotides (cyan) are presented for both the strands. These values in the central A,T-rich core are consistently higher than those in the flanking region.

Figure 4.

Overall ionic environment and crystal packing. (A) The duplex is drawn in stick representation. The calcium ions surrounding the duplex are drawn as black and gray (symmetry-mates) spheres. A total of 20 ions including the symmetry-related are found on the surface of a parent molecule. Calcium ions located on the 3-fold axes occur in pairs. Of the 12 independent calcium ions, eight lie on 3-fold axes and the remaining four are found in the minor groove. These calcium ions serve dual role in neutralizing the negative charge on the duplex as well as in stabilizing crystal packing. (B) Stereo view of three columns of stacked duplexes in the crystal. A crystallographic 3-fold axis relates the three columns. The duplexes are drawn as ribbons with the parent molecule in cyan and its 3-fold symmetry-related duplexes in green and pink.

The stacked duplexes at the top and bottom are in gray. The ionic environment surrounding the parent duplex is shown with calcium ions drawn as orange spheres that belong to the parent molecule, while the gray spheres are symmetry-related ions. The black circles at one of the three columns indicate the missing phosphate groups that connect one duplex with another.

Figure 5.

Schematic representation of minor-groove hydration. The bases are depicted as gray (A-T base pairs) and black (C-G base pairs) rectangles. The deoxyribose sugars connect the phosphate groups labeled as P to the bases. The characteristic spine of water molecules found in the narrow central A,T-rich region is shown as spheres colored in cyan. Hydrogen bonds connecting water molecules in the spine are shown as thick broken lines. Two additional water molecules are seen at both ends of the spine resulting in two water molecules mediating cross strand interaction between O2 and N3 atoms instead of one seen for the water molecules seen in the interior of the spine. Thin broken lines indicate hydrogen bonds between water molecules and DNA atoms. A specific pattern of water molecules that is found in the central region is

interrupted at both ends of the duplex. Four calcium ions drawn as spheres colored in orange are bound at the ends of the minor groove. Two linear arrays of water molecules shown as blue spheres span the groove on either side of the spine. These water molecules mediate interactions between DNA atoms and calcium ions.

Figure 6.

Hydration and calcium ions in the minor groove. (A) DNA duplex is shown in stick representation with the consensus DNA site in purple and the remaining residues in cyan. Water molecules spanning the floor of the minor groove are drawn as violet spheres. The calcium ions are displayed as orange spheres. (B) View is 180° rotated about the helix axis. The floor of the minor groove is filled with water molecules and calcium ions. There is a split in the characteristic spine of hydration into two linearly ordered water molecules that are coordinated with calcium ions at both ends of the groove.

Figure 7.

Stereo views of environment of calcium binding. The parent molecule is in green and the symmetry-related molecules are in different colors. Water molecules and phosphate oxygens that are bonded to Ca^{2+} ions are drawn as spheres. The immediate surroundings of the calcium coordination are displayed. (A) Vicinity of outer calcium ion in the minor groove. It is surrounded by water molecules except for one non-bridging phosphate oxygen. (B) The surroundings of inner calcium ion in the minor groove. All of its ligands are water molecules. (C) A calcium ion pair located on 3-fold axis links three symmetry related duplexes laterally. (D) A calcium ion pair at a terminus of a duplex connects a set of three duplexes to another set of three duplexes forming infinite columns of helices.

Figure 8. Minor-groove width variation in dsx^A . Surface representation of DNA duplex with the sugar-phosphate backbone and the bases shown in green and white, respectively. (A) The central A-T-rich region assumes a minimum value. (B) This view (model in panel A rotated by 180° about the helix axis) highlights the symmetrical widening of the minor groove on either side of the inner A-T-rich portion.

Figure 9.

Plot of the minor-groove width. The values for groove width corresponds to P---P distances between the appropriate phosphate groups on the two strands minus 5.8 Å for the radii of phosphate groups. The groove is narrowest in the central A-T-rich region and widens on either side.

Table 1Cognate DNA Sub-sites (5'→3') of transcription factors[†] found in the present structure

protein	DNA target site
A. The DSX and the shared SRY binding sites	
Present study (16 base-pair duplex)	5'-ACTACAATGTTGCAAT-3'
DSX ^a	ACAATGTT
SRY ^b	ACAAT
B. Sub-sites (4- or more base-pairs) found within the 16-bp duplex	
FOXJ2	ACAAT
Sox-5/Sox-9 ^c	ACAAT
Mat1-Mc	ACAAT
HNF-3β ^d	AATGTT
C/EBP ^e	AATGTTGCAAT
CHOP-C/EBPα ^f	TGCAAT
FOXD3	ATGTT
Oct-1 ^g	CAATGT
MIF-1	GTTGC
VBP	TACA
RFX1	GTTGCAAT
HLF	GCAAT
MATa1 ^h	ATGT
Elf-1	CTACA
Athb-1	TTGCA
AbaA	AATGT
c-Myb ⁱ	GTTG
C. Sub-sites spanning the pseudo-continuous helical section of the duplexes in the crystal	
v-Myb	AATA (...GCAAT-3' adjoining 5'-ACTC...)
GATA-1 ^j	ATACT (...GCAAT-3' adjoining 5'-ACTAC...)
CDP CR3	TATTG (...CGTTA-5' adjoining 3'-TGAT...)

[†] SRY, sex-determining region Y protein; FOXJ2, fork head box J 2; Sox, SRY-related HMG-box gene; Mat1-Mc, M-box interacting with Mat1-Mc; HNF-3β, Hepatocyte Nuclear Factor 3β; C/EBP, C/EBP binding site; CHOP-C/EBPα, heterodimers of CHOP and C/EBPα; FOXD3, fork head box D3; Oct-1, octamer-binding factor 1; MIF-1, MIBP-1 / RFX1 complex; VBP, PAR-type chicken vitellogenin promoter-binding protein; RFX1, X-box binding protein RFX1; HLF, hepatic leukemia factor; MATa1, mating factor a1; Elf-1, Elf-1 binding site; Athb-1, Arabidopsis thaliana homeo box protein 1; AbaA, AbaA binding site; c-Myb, c-Myb binding site; v-Myb, viral myb binding site; GATA-1, GATA-binding factor 1; CDP CR3, cut-like homeodomain protein. The availability of NMR or crystal structures (protein or its homolog or DNA target site or protein-DNA complex) of a specific transcription factor is indicated by a superscript letter and the respective PDB codes are listed below:

^a 1lpv;^b 1lwa, 1j46;^c 1i11;^d 2hfh;^e 2e43;^f 1nwq;^g 1hf0;^h 1f43;ⁱ 1bdz, 1idy;^j 3gat

Table 2

Base-pair helical parameters

Base-pair step	Slide (Å)	Roll (°)	Twist (°)	Base-pair	Buckle (°)	Propeller (°)	Opening (°)
A3-C4	-0.7	5.2	28.2	A3-T36	4.2	-7.1	3.5
C4-T5	-0.3	0.7	29.3	C4-G35	-2.0	-14.1	-0.1
T5-A6	1.2	4.2	41.5	T5-A34	2.1	-19.8	2.3
A6-C7	-0.2	4.6	21.8	A6-T33	-5.4	-7.1	-1.8
C7-A8	2.2	-3.6	47.9	C7-G32	6.8	-2.8	2.2
A8-A9	0.1	3.6	31.9	A8-T31	12.8	-15.7	2.7
A9-T10	-0.7	-1.5	32.3	A9-T30	9.2	-15.6	1.4
T10-G11	-0.5	8.5	37.6	T10-A29	7.4	-15.8	4.2
G11-T12	-0.7	-0.9	33.0	G11-C28	-12.6	-13.7	3.6
T12-T13	0.1	2.6	31.7	T12-A27	-14.4	-16.2	2.0
T13-G14	2.1	-6.0	49.1	T13-A26	-15.1	-12.5	4.1
G14-C15	-0.2	4.7	21.5	G14-C25	-8.9	-4.0	1.4
C15-A16	1.1	1.7	42.0	C15-G24	5.5	-3.8	0.4
A16-A17	-0.5	3.5	29.5	A16-T23	2.9	-16.0	1.2
A17-T18	-0.8	3.0	27.8	A17-T22	-1.2	-12.6	-1.3
Average	0.2	2.0	33.7	T18-A21	-2.0	-7.8	2.7
S.D	1.0	3.7	8.4		-0.7	-11.5	1.8
					8.6	5.3	1.8

Table 3

Comparison of base-pair helical parameters in 5'-ACAA-3' subsite of dsx^A and *MIS* segment

Base-pair step	Rise (Å)	Roll (°)		Twist (°)		Base-pair		Buckle (°)		Propeller (°)		Opening (°)		<i>MIS</i>
		<i>dsx^A</i>	<i>MIS</i>	<i>dsx^A</i>	<i>MIS</i>	<i>dsx^A</i>	<i>MIS</i>	<i>dsx^A</i>	<i>MIS</i>	<i>dsx^A</i>	<i>MIS</i>	<i>dsx^A</i>	<i>MIS</i>	
A-C	3.1	3.7	3.7	21.8	32.8A-T	-5.4	2.1	-7.1	-13.9	-1.8	-2.8	-1.7	-1.3	-2.8
C-A	3.3	3.9	3.9	47.9	37.5C-G	6.8	-2.0	-2.8	-5.7	2.2	-2.6	2.2	-5.7	-2.6
A-A*	3.4	3.3	3.3	31.9	22.2A-T	12.8	-15.1	-15.7	-15.9	2.7	0.0	2.7	-15.9	0.0
A-T	3.3	4.5	4.5	32.3	27.6A-T	9.2	14.7	-15.6	7.7	1.4	-1.4	1.4	7.7	-1.4
Average	3.3	3.8	3.8	33.5	30.0	5.8	-0.1	-10.3	-6.9	1.1	-1.7	1.1	-6.9	-1.7
S.D	0.1	0.5	0.5	10.8	6.6	7.9	12.3	6.4	10.7	2.0	1.3	2.0	10.7	1.3

* A-T step is A-A in *MIS* segment. DNA sites are designated dsx^A (target site for Doublesex) and *MIS* (target site for SRY-related HMG boxes).

Table 4

Statistics for data collection, phasing, and refinement

		<i>Data collection and MAD phasing</i>			
Space group	R3				
Unit cell (Å)	a=b=38.74, c=161.33; γ=120o (hexagonal setting)				
V _m (Å ³ /Da)	2.2 (one duplex / asymmetric unit; 54 % solvent content)				
Data set	Inflexion (λ _i)	Peak (λ _p)	Remote (λ _r)	Native	
Wavelength (Å)	0.9202	0.9200	0.9135	0.9186	
Resolution (Å)	30-2.2	30-2.2	30-2.2	20-1.6	
Observed reflections	47,253	46,501	29,713	82,087	
Unique reflections	4,638	5,669	5,778	11,813	
Completeness (%)	95.2	94.8	96.6	99.7	
Average I/σ ^a	26.8 (4.3)	26.1 (4.1)	23.9 (3.7)	10.1 (6.7)	
R _{sym} (%) ^{a,b}	3.8 (26)	4.0 (26)	3.7 (24)	4.1 (17)	
Mean figure of merit (FOM) ^c		0.76			
		<i>Native data refinement statistics</i>			
Resolution range (Å)	8-1.6				
R _{cryst} (%) (F > 4σ _F) ^d	23.1				
R _{free} (%) (F > 4σ _F)	27.4				
R (%) (F > 4σ _F)	22.1				
Reflections (F > 4σ _F ; completeness)					
Working set	10,478 (88.5 %)				
Test set	1,189 (10 %)				
All (Working + Test)	11,667 (98.5 %)				
Number of non-H atoms					
DNA	658				
Solvent	197				
Calcium ions	12				
Average B values (Å ²)					
DNA	15.2				
Solvent	26.9				
Calcium ions	15.6				

^a Values for the highest resolution shell are given within parentheses.^b $R_{\text{sym}} = \sum_{\text{hkl}} \sum_i |I_i(\text{hkl}) - \langle I(\text{hkl}) \rangle| / \sum_{\text{hkl}} \sum_i I_i(\text{hkl})$. Friedel-mates were collected at wavelengths λ_i and λ_p .^c MAD phases from Bromine derivative were computed using data between 30.0 - 2.2 Å resolution. FOM = $\langle \cos(\Delta\alpha) \rangle$, in which $\Delta\alpha$ indicates the error in phase angle of a reflection.^d $R_{\text{cryst}} = \sum_{\text{hkl}} |F_{\text{obs}}(\text{hkl}) - F_{\text{calc}}(\text{hkl})| / \sum_{\text{hkl}} F_{\text{obs}}(\text{hkl})$.